หนังสือเรียนรายวิชาเพิ่มเติม

ชั้นมัธยมศึกษาปีที่ ๖

คณิตศาสตร์

เล่ม ๒

ตามผลการเรียนรู้
กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. ๒๕๖๐)
ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑



หนังสือเรียน

รายวิชาเพิ่มเติม

คณิตศาสตร์

ชั้น

มัธยมศึกษาปีที่ ๖ เล่ม ๒

ตามผลการเรียนรู้
กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. ๒๕๖๐)
ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑

จัดทำโดย
สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ

จัดทำเป็นฉบับ e-book ครั้งที่ ๑ พ.ศ. ๒๕๖๓

คำชี้แจง

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) ได้จัดทำตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช ๒๕๕๑ โดยมีจุดเน้นเพื่อต้องการพัฒนาผู้เรียนให้มีความรู้ความสามารถที่ทัดเทียมกับนานาชาติ ได้เรียนรู้คณิตศาสตร์ที่เชื่อมโยงความรู้กับกระบวนการคิด ใช้กระบวนการสืบเสาะหาความรู้และแก้ปัญหาที่หลากหลาย มีการทำกิจกรรมด้วยการลงมือปฏิบัติ เพื่อให้ผู้เรียนได้ใช้ทักษะกระบวนการทางคณิตศาสตร์และทักษะแห่งศตวรรษที่ ๒๑ ซึ่งในปีการศึกษา ๒๕๖๑ เป็นต้นไป โรงเรียนจะต้องใช้หลักสูตรกลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. ๒๕๖๐) สสวท. จึงได้จัดทำหนังสือเรียนที่เป็นไปตามมาตรฐานหลักสูตรเพื่อให้โรงเรียนได้ใช้สำหรับจัดการเรียนการสอนในชั้นเรียน

หนังสือเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ ๖ เล่ม ๒ นี้ มีผลการเรียนรู้และสาระการเรียนรู้เพิ่มเติมที่ครอบคลุมเนื้อหาบางส่วนที่ปรากฏตามตัวชี้วัดรายวิชาพื้นฐานคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ ๖ โดยเมื่อผู้เรียนเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ ทั้ง ๖ เล่ม ครบทุกชั้นปีในระดับชั้นมัธยมศึกษาปีที่ ๔ – ๖ แล้วจะสามารถบรรลุผลสัมฤทธิ์ตามตัวชี้วัดของรายวิชาพื้นฐานคณิตศาสตร์ได้ และในขณะเดียวกันก็สามารถต่อยอดเนื้อหาจากรายวิชาพื้นฐานไปสู่เนื้อหาในรายวิชาเพิ่มเติมได้โดยไม่ต้องเสียเวลาเรียนซ้ำซ้อน ทั้งนี้ หนังสือเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ เล่ม ๒ นี้ มีเนื้อหาที่จำเป็นที่ต้องเรียน ประกอบด้วยเรื่องความหมายของสถิติศาสตร์และข้อมูล การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพ การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ และตัวแปรสุ่มและการแจกแจงความน่าจะเป็น ซึ่งเป็นพื้นฐานที่สำคัญและเพียงพอสำหรับการศึกษาต่อในระดับอุดมศึกษาในด้านวิทยาศาสตร์ หรือประกอบอาชีพในสาขาที่ใช้วิทยาศาสตร์เป็นฐาน เช่น แพทย์ ทันตแพทย์ สัตวแพทย์ เทคโนโลยีชีวภาพ วิศวกรรม สถาปัตยกรรม เศรษฐศาสตร์ พาณิชยศาสตร์ ฯลฯ โดยเน้นกระบวนการคิดวิเคราะห์และการแก้ปัญหา เชื่อมโยงความรู้สู่การนำไปใช้ในชีวิตจริง ตลอดจนมีกิจกรรมที่ส่งเสริมให้ผู้เรียนทำงานร่วมกับผู้อื่นได้อย่างมีประสิทธิภาพ รวมทั้งสามารถใช้คณิตศาสตร์เป็นเครื่องมือในการเรียนรู้วิทยาศาสตร์ เทคโนโลยี และศาสตร์อื่น ๆ ที่เกี่ยวข้อง อันเป็นรากฐานในการพัฒนาทรัพยากรบุคคลของชาติให้มีคุณภาพและพัฒนาเศรษฐกิจของประเทศ

การจัดทำหนังสือเรียนเล่มนี้ ได้รับความร่วมมืออย่างดียิ่งจากผู้ทรงคุณวุฒิ นักวิชาการอิสระ คณาจารย์ทั้งหลาย รวมทั้งครูผู้สอน นักวิชาการ จากสถาบันและสถานศึกษาทั้งภาครัฐและเอกชน จึงขอขอบคุณทุกท่านไว้ ณ ที่นี้ หากมีข้อเสนอแนะใดที่จะทำให้หนังสือเรียนเล่มนี้มีความสมบูรณ์ยิ่งขึ้น โปรดแจ้ง สสวท. ทราบด้วย จะขอบคุณยิ่ง

(ศาสตราจารย์ชูกิจ ลิมปิจำนงค์)
ผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
กระทรวงศึกษาธิการ

คำอธิบายรายวิชา

หนังสือเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ **ชั้นมัธยมศึกษาปีที่ ๖ เล่ม ๒**

กลุ่มสาระการเรียนรู้คณิตศาสตร์

ภาคเรียนที่ ๒ เวลา ๑๐๐ ชั่วโมง **จำนวน ๒.๕ หน่วยกิต**

ศึกษา พร้อมทั้งฝึกทักษะและกระบวนการทางคณิตศาสตร์ในเนื้อหาของสาระ ดังนี้

ความหมายของสถิติศาสตร์และข้อมูล สถิติศาสตร์ คำสำคัญในสถิติศาสตร์ ประเภทของข้อมูล สถิติศาสตร์เชิงพรรณนาและสถิติศาสตร์เชิงอนุมาน

การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพ การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่ การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่ การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ ค่าวัดทางสถิติ

ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น ความหมายและชนิดของตัวแปรสุ่ม การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง การแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง

โดยจัดประสบการณ์ให้ผู้เรียนได้พัฒนาทักษะและกระบวนการทางคณิตศาสตร์ อันได้แก่ การแก้ปัญหา การสื่อสารและการสื่อความหมายทางคณิตศาสตร์ การเชื่อมโยง การให้เหตุผล และการคิดสร้างสรรค์

การใช้สื่อ อุปกรณ์ เทคโนโลยี และแหล่งข้อมูล และนำประสบการณ์ ตลอดจนทักษะและกระบวนการที่ได้ ไปใช้ในการเรียนรู้สิ่งต่าง ๆ และใช้ในชีวิตประจำวันอย่างสร้างสรรค์ รวมทั้งเห็นคุณค่าและมีเจตคติที่ดีต่อคณิตศาสตร์ สามารถทำงานอย่างเป็นระบบ มีความรอบคอบ และมีวิจารณญาณ

การวัดผลประเมินผล ใช้วิธีการที่หลากหลายตามสภาพความเป็นจริงให้สอดคล้องกับเนื้อหาและทักษะที่ต้องการวัด

ผลการเรียนรู้

๑. เข้าใจและใช้ความรู้ทางสถิติในการนำเสนอข้อมูล และแปลความหมายของค่าสถิติเพื่อประกอบการตัดสินใจ

๒. หาความน่าจะเป็นของเหตุการณ์ที่เกิดจากตัวแปรสุ่มที่มีการแจกแจงเอกรูป การแจกแจงทวินาม และการแจกแจงปกติ และนำไปใช้ในการแก้ปัญหา

รวมทั้งหมด ๒ ผลการเรียนรู้

แนะนำการใช้หนังสือเรียน

ส่วนนี้เป็นส่วนแนะนำโครงสร้างของหนังสือเรียนเพื่อการใช้หนังสือเรียนอย่างมีประสิทธิภาพ

ในหนังสือเล่มนี้จะแบ่งบทเรียนเป็น 4 บท โดยแต่ละบทจะมีส่วนประกอบ ดังนี้

“
....
....
”

ส่วนนำของบท

เกริ่นนำบทด้วยข้อมูลที่น่าสนใจ รวมถึงการนำไปใช้ในชีวิตจริง เพื่อกระตุ้นให้นักเรียนตระหนักถึงความสำคัญของการเรียนเนื้อหาในบท

จุดมุ่งหมาย

เป้าหมายที่นักเรียนควรไปถึงหลังจากเรียนจบบทนี้

จุดมุ่งหมาย

ความรู้ก่อนหน้า

ความรู้ที่นักเรียนจำเป็นต้องมีก่อนที่จะเรียนบทนี้

ความรู้ก่อนหน้า

เสริมสมอง

เกร็ดความรู้หลากหลายรูปแบบเพื่อกระตุ้นความสนใจของนักเรียน เช่น ประวัตินักคณิตศาสตร์ ตัวอย่างการนำเนื้อหาคณิตศาสตร์ไปใช้ในชีวิตจริง การเชื่อมโยงเนื้อหาในบทกับวิชาอื่น

เสริมสมอง

กิจกรรม

กิจกรรมที่นักเรียนสามารถศึกษาเพิ่มเติมได้ด้วยตนเอง เพื่อช่วยพัฒนาทักษะการเรียนรู้และนวัตกรรม (learning and innovation skills) ที่จำเป็นสำหรับศตวรรษที่ 21 อันได้แก่ การคิดสร้างสรรค์และนวัตกรรม (creativity and innovation) การคิดแบบมีวิจารณญาณและการแก้ปัญหา (critical thinking and problem solving) การสื่อสาร (communication) และการร่วมมือ (collaboration)

กิจกรรม

เทคโนโลยี

โจทย์ที่มีไอคอนนี้สามารถใช้เทคโนโลยีช่วยในการคิดได้ โดยอาจใช้เครื่องคิดเลข โปรแกรมสำเร็จรูปในคอมพิวเตอร์ โปรแกรมประยุกต์ในโทรศัพท์สมาร์ตโฟน การค้นหาข้อมูลทางอินเทอร์เน็ต และอื่น ๆ ตามความเหมาะสม

โจทย์ท้าทาย

โจทย์ที่มีไอคอนนี้เป็นโจทย์ที่ค่อนข้างยากและซับซ้อนกว่าโจทย์แบบฝึกหัดทั่วไป เพื่อท้าทายความสามารถของนักเรียนที่ต้องการพัฒนาทักษะเพิ่มเติมนอกเหนือจากการเรียนตามจุดมุ่งหมายของบท

แบบฝึกหัด

โจทย์เพื่อตรวจสอบความรู้ระหว่างเรียน มีวัตถุประสงค์เพื่อพัฒนาการเรียนรู้ของนักเรียน ช่วยให้นักเรียนสามารถตรวจสอบความรู้ความเข้าใจของตนเองได้

แบบฝึกหัด

แบบฝึกหัดท้ายบท

แบบฝึกหัดท้ายบทแบ่งประเภทได้เป็น

- โจทย์เพื่อตรวจสอบความรู้หลังเรียนจบบท มีวัตถุประสงค์เพื่อวัดความรู้ความเข้าใจของนักเรียนตามจุดมุ่งหมายของบท โดยจะมีแถบสี
- โจทย์ท้าทาย
- โจทย์เพื่อฝึกทักษะ ที่มีความน่าสนใจ โดยจะไม่มีแถบสีและไม่มีไอคอนหน้าเลขข้อ

แบบฝึกหัดท้ายบท

1. ....
....

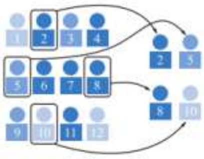
2. ....
....

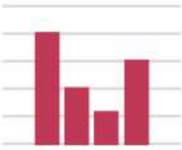
3. ....
....

บทที่ 1 จะใช้สี บทที่ 2 จะใช้สี บทที่ 3 จะใช้สี บทที่ 4 จะใช้สี

สารบัญ

บทที่ 1 – 2

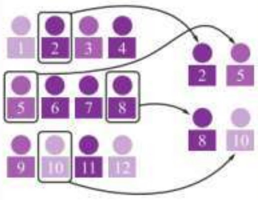




สารบัญ บทที่ 3 – 4

บทที่	เนื้อหา	หน้า

บทที่ 1 | ความหมายของสถิติศาสตร์และข้อมูล



จุดมุ่งหมาย

1. บอกความหมายและประโยชน์ของสถิติศาสตร์ พร้อมทั้งสามารถอธิบายได้ว่าการนำเสนอข้อมูลที่พบเห็นในชีวิตประจำวัน มีความน่าเชื่อถือมากน้อยเพียงใด ก่อให้เกิดความเข้าใจคลาดเคลื่อนหรือไม่
2. ระบุประชากร ตัวอย่าง ตัวแปร ข้อมูล พารามิเตอร์ และค่าสถิติจากสถานการณ์ที่กำหนด
3. จำแนกประเภทของข้อมูลตามแหล่งที่มาของข้อมูล ระยะเวลาที่จัดเก็บหรือลักษณะของข้อมูล
4. ระบุได้ว่าสถานการณ์ที่กำหนดใช้วิธีการของสถิติศาสตร์เชิงพรรณนาหรือสถิติศาสตร์เชิงอนุมาน

บทที่ 1

ความหมายของสถิติศาสตร์และข้อมูล

“

ข้อมูลที่เกิดขึ้นในปัจจุบันมีหลายรูปแบบ และสามารถนำมาใช้ให้เกิดประโยชน์ในด้านต่าง ๆ เช่น ในการดำเนินธุรกิจ ผู้ประกอบการอาจมีความต้องการที่จะปรับปรุงคุณภาพของสินค้าที่ผลิตขึ้น หรือหาแนวทางใหม่ ๆ ในการให้บริการแก่ลูกค้า ดังนั้น ผู้ประกอบการจึงต้องมีการเก็บข้อมูลต่าง ๆ ที่เกี่ยวกับธุรกิจของตน ไม่ว่าจะเป็นการผลิตสินค้า การเก็บรักษาสินค้าและการขนส่งสินค้า การจ้างบุคลากรในส่วนต่าง ๆ รายรับและรายจ่ายในแต่ละประเภท และอื่น ๆ หากพิจารณาการประกอบกิจการธุรกิจร้านขายเครื่องดื่มและขนมหวาน ผู้ประกอบการจะต้องเก็บข้อมูลปริมาณวัตถุดิบที่ต้องใช้ในการผลิต เช่น เมล็ดกาแฟ นม น้ำตาล แป้งสาลี เนย ผลไม้หรือธัญพืชที่ใช้ในการทำขนม ราคาและเวลาที่ใช้ในการสั่งซื้อวัตถุดิบแต่ละชนิด แรงงานหรือทรัพยากรอื่น ๆ ที่ต้องใช้ในการผลิตเครื่องดื่มและขนมแต่ละชนิด ปริมาณลูกค้าที่เข้ามาใช้บริการในแต่ละช่วงเวลาของวัน สัปดาห์ เดือน หรือช่วงเทศกาลต่าง ๆ รวมทั้งระยะเวลาที่ลูกค้าเข้ามาใช้บริการในร้าน นอกจากนี้ ผู้ประกอบการอาจสอบถามความพึงพอใจของลูกค้า โดยเก็บข้อมูลเกี่ยวกับเพศ อายุ เครื่องดื่มหรือขนมที่ลูกค้านิยม ระดับความพึงพอใจต่อการให้บริการของพนักงาน รวมทั้งความเห็นอื่น ๆ เกี่ยวกับร้าน แล้วนำข้อมูลเหล่านี้มาวิเคราะห์และใช้ประกอบการพิจารณาเพื่อออกแบบผลิตภัณฑ์ใหม่ให้ตอบสนองความต้องการของลูกค้าได้มากขึ้น หรือปรับปรุงการให้บริการ การตัดสินใจทางธุรกิจบนพื้นฐานของการวิเคราะห์ข้อมูลดังตัวอย่างข้างต้นเป็นปัจจัยสำคัญที่ส่งผลต่อความสำเร็จของธุรกิจต่าง ๆ ในปัจจุบัน และความรู้ที่อยู่เบื้องหลังการวิเคราะห์ข้อมูลดังกล่าวก็คือความรู้ทางด้านสถิติศาสตร์

”

ความรู้ก่อนหน้า

- ความรู้เกี่ยวกับสถิติศาสตร์ในระดับมัธยมศึกษาตอนต้น

ipst.me/10676

1.1 สถิติศาสตร์

สถิติศาสตร์ (statistics) หมายถึง วิชาที่ว่าด้วยการเก็บรวบรวมข้อมูล วิเคราะห์ข้อมูล และสรุปผลจากข้อมูลที่เกี่ยวข้อง เพื่อนำมาตอบคำถาม อธิบายปรากฏการณ์หรือประเด็นที่สนใจ

ประโยชน์ของสถิติศาสตร์ในด้านต่าง ๆ เช่น

ด้าน	ตัวอย่างการนำไปใช้ประโยชน์
การศึกษา	รัฐบาลสามารถนำข้อมูลเกี่ยวกับประชากรก่อนวัยเรียนและวัยเรียน บุคลากรทางการศึกษา ปริมาณการผลิตและพัฒนาครูในแต่ละสาขาวิชา จำนวนสถานศึกษา และค่าใช้จ่ายในแต่ละระดับการศึกษา มาวิเคราะห์เพื่อกำหนดนโยบายและวางแผนพัฒนาการศึกษาและการกระจายโอกาสทางการศึกษาให้กับเด็กและเยาวชนในระดับการศึกษาขั้นพื้นฐาน
การเกษตร	รัฐบาลสามารถนำข้อมูลเกี่ยวกับครัวเรือนที่ทำการเกษตร พื้นที่การเพาะปลูก ผลิตผลทางการเกษตร ราคาสินค้าเกษตรกรรม เครื่องมือเครื่องใช้ทางการเกษตร แหล่งน้ำและการชลประทาน และความต้องการสินค้าทางการเกษตรของผู้บริโภค มาวิเคราะห์เพื่อให้คำแนะนำในการวางแผนการเพาะปลูกกับเกษตรกรในฤดูกาลถัดไปได้

ด้าน	ตัวอย่างการนำไปใช้ประโยชน์
การผลิต	การวางแผนผลิตสินค้าออกจำหน่ายควรเก็บข้อมูลและวิเคราะห์ว่าสินค้านั้นเป็นที่ต้องการของตลาดหรือไม่ ผู้ผลิตมีกำลังการผลิตเท่าใด ควรจำหน่ายสินค้าที่ใดและทำอย่างไรผลผลิตจะมีอย่างต่อเนื่องและได้สินค้าที่มีคุณภาพ
การควบคุมคุณภาพสินค้า	การควบคุมคุณภาพสินค้าที่ผลิตให้มีมาตรฐานตามที่กำหนด โดยเฉพาะอย่างยิ่งในยุคที่มีการแข่งขันสูง การรักษามาตรฐานของสินค้ายิ่งมีความสำคัญมากยิ่งขึ้น อย่างไรก็ตาม การตรวจสอบคุณภาพของสินค้าที่ผลิตทุกชิ้นก่อนส่งออกจำหน่ายนับเป็นภาระที่มากเกินกว่าที่จะกระทำได้ ในการควบคุมคุณภาพสินค้าที่ผลิตให้เป็นไปตามมาตรฐานที่กำหนดโดยไม่ต้องตรวจสอบคุณภาพของสินค้าที่ผลิตทุกชิ้นเป็นสิ่งที่สามารถกระทำได้ โดยกำหนดวิธีการตามขั้นตอนของกระบวนการทางสถิติศาสตร์ กล่าวคือ ในขั้นตอนการเก็บข้อมูล สถิติศาสตร์จะกำหนดวิธีการเก็บรวบรวมข้อมูลที่เหมาะสมว่าควรเป็นการเลือกตัวอย่างโดยการสุ่มตัวอย่างสินค้าจำนวนหนึ่งจากจำนวนสินค้าทั้งหมดที่ผลิตในแต่ละรุ่น โดยสามารถกำหนดขนาดตัวอย่างที่จะเลือกและสามารถใช้วิธีการวิเคราะห์ซึ่งสร้างเกณฑ์การพิจารณาว่า หากมีสินค้าที่ชำรุดในสินค้าตัวอย่างทั้งหมดที่เลือกมาจำนวนไม่เกินเท่าใดแล้วจะถือว่าสินค้ารุ่นนั้นได้มาตรฐานตามที่กำหนด
ผู้บริโภค	การตัดสินใจเลือกซื้อสินค้าสามารถนำสถิติศาสตร์มาใช้ โดยเก็บข้อมูลเกี่ยวกับสินค้าที่ต้องการ แล้วนำข้อมูลที่ได้ไปวิเคราะห์หรือเปรียบเทียบข้อมูลจากร้านค้าต่าง ๆ ที่สำรวจเพื่อนำมาตัดสินใจว่าจะซื้อสินค้านั้นหรือไม่
การพยากรณ์	การพยากรณ์เกี่ยวกับฝนโดยอาศัยข้อมูลจากกรมอุตุนิยมวิทยา ซึ่งใช้ข้อมูลทั้งในอดีตและปัจจุบันตลอดจนวิธีวิเคราะห์ทางสถิติศาสตร์เข้ามาช่วย ทำให้สามารถคาดเดาสถานการณ์ล่วงหน้าได้ว่าฝนจะตกหนักหรือไม่
ชีวิตประจำวัน	จดบันทึกรายรับรายจ่ายในแต่ละเดือนเพื่อใช้ในการวิเคราะห์พฤติกรรมการใช้จ่ายซึ่งจะช่วยให้สามารถปรับวิธีการใช้จ่ายเงินและควบคุมค่าใช้จ่ายได้อย่างมีประสิทธิภาพมากยิ่งขึ้น

จะเห็นว่าสถิติศาสตร์มีความสำคัญอย่างมาก โดยเฉพาะอย่างยิ่งในปัจจุบันซึ่งมีข้อมูลข่าวสารมากมายอยู่รอบตัวเรา ผู้วิเคราะห์ข้อมูลจำเป็นต้องมีความรู้ความเข้าใจเกี่ยวกับเครื่องมือและกระบวนการทางสถิติศาสตร์ที่เหมาะสมกับสถานการณ์ต่าง ๆ และสามารถแปลความหมายโดยไม่ก่อให้เกิดความเสียหายหรือความเข้าใจคลาดเคลื่อนในภายหลัง นอกจากนี้ ผู้ใช้ข้อมูลยังต้องมีความรู้ทางสถิติศาสตร์ที่จำเป็น เพื่อให้สามารถสังเกตเห็นข้อมูลที่เป็นเท็จหรือทำให้เข้าใจผิด และตระหนักถึงข้อมูลที่บิดเบือนไปจากความเป็นจริง รวมทั้งสามารถวิเคราะห์ได้ว่าค่าสถิติ แผนภูมิ กราฟ หรือการนำเสนอข้อมูลต่าง ๆ มีความน่าเชื่อถือมากน้อยเพียงใด ก่อให้เกิดความเข้าใจคลาดเคลื่อนหรือไม่ ตัวอย่างเช่น ในการวิเคราะห์ผลการสำรวจความคิดเห็นของนักเรียนที่มีต่อการไปโรงเรียน ดังนี้

ผลสำรวจความคิดเห็นของนักเรียนที่มีต่อการไปโรงเรียน

	ร้อยละ
รู้สึกดี	62.7
รู้สึกเฉย ๆ	29.8
รู้สึกไม่ดี	3.5

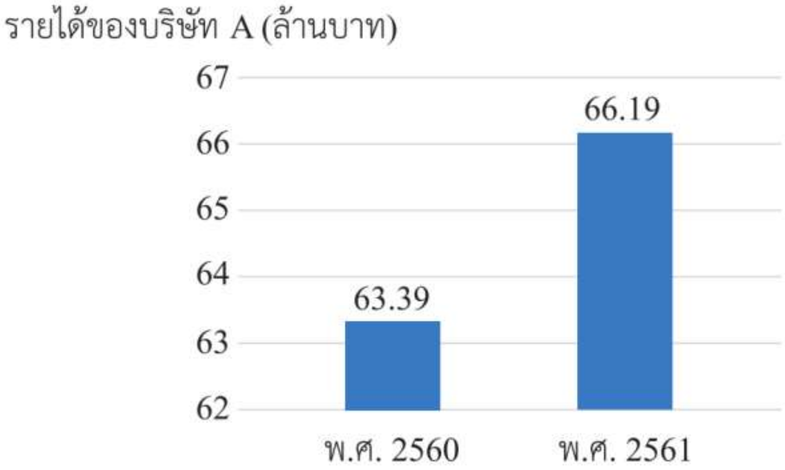
เมื่อพิจารณาอย่างผิวเผิน ผลสำรวจดังกล่าวอาจดูปกติ แต่ถ้าพิจารณาให้รอบคอบจะพบว่า ข้อมูลดังกล่าวไม่ได้บอกรายละเอียดในการสำรวจ เช่น จำนวนนักเรียนทั้งหมดที่สำรวจ และเมื่อนำร้อยละของความคิดเห็นทั้งหมดมารวมกันก็ไม่เท่ากับ 100 ดังนั้นจึงยังไม่สามารถสรุปได้ว่า ข้อมูลดังกล่าวน่าเชื่อถือหรือไม่

ค่าเฉลี่ยเลขคณิตเป็นค่าที่มักพบในการรายงานทั่วไป เช่น บริษัท A รายงานว่า พนักงานของบริษัทมีเงินเดือนเฉลี่ย 40,000 บาท หลายคนอาจตีความว่า พนักงานส่วนใหญ่ของบริษัท A มีเงินเดือนประมาณ 40,000 บาท แต่เมื่อพิจารณาข้อมูลเงินเดือนของพนักงานทั้งหมดตามที่แสดงในตาราง จะพบว่ามีพนักงานเพียง 1 คนเท่านั้นที่มีเงินเดือนเท่ากับค่าเฉลี่ย ในขณะที่พนักงานส่วนใหญ่มีเงินเดือนอยู่ในช่วง 15,000 – 25,000 บาท ซึ่งเงินเดือนน้อยกว่าค่าเฉลี่ยประมาณครึ่งหนึ่ง และมีพนักงาน 4 คนเท่านั้นที่มีเงินเดือนเกินค่าเฉลี่ย การนำเสนอข้อมูลด้วยค่าเฉลี่ยเลขคณิตจึงไม่เหมาะสมในการอธิบายภาพรวมของตัวอย่างนี้ เนื่องจากอาจก่อให้เกิดความเข้าใจคลาดเคลื่อนได้ โดยเฉพาะในกรณีที่ขาดความรู้ความเข้าใจเกี่ยวกับสถิติศาสตร์

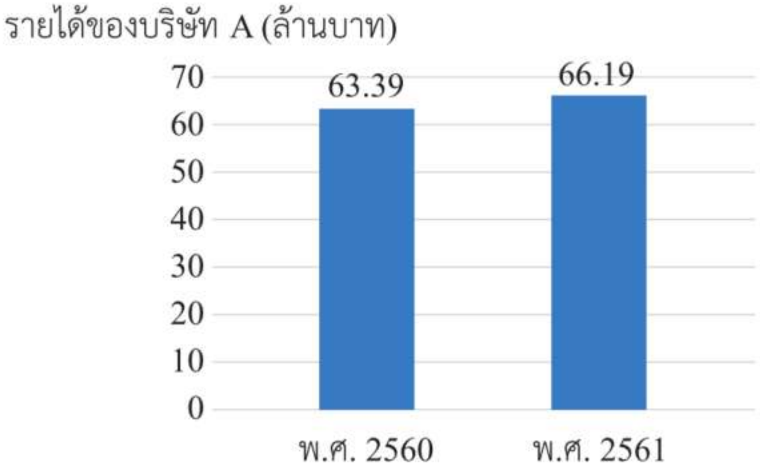
เงินเดือนของพนักงานทั้งหมดในบริษัท A เรียงลำดับจากน้อยไปมาก

เงินเดือน (บาท)	จำนวนพนักงาน (คน)
15,000	5
20,000	4
22,000	4
25,000	3
40,000	1
91,000	2
120,000	1
180,000	1

ตัวอย่างการนำเสนอข้อมูลที่อาจก่อให้เกิดความเข้าใจคลาดเคลื่อนได้ในรูปแบบอื่น ๆ เช่น การนำเสนอข้อมูลด้วยแผนภูมิแท่งดังต่อไปนี้



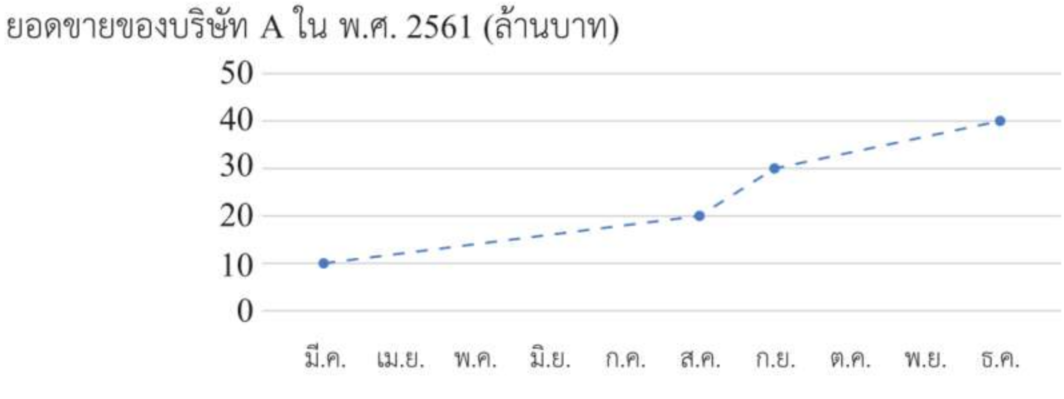
จากแผนภูมิแท่ง ถ้าพิจารณาจากความสูงของแท่งซึ่งแทนรายได้ จะเห็นว่ารายได้ของบริษัท A ใน พ.ศ. 2561 มีค่าประมาณสามเท่าของรายได้ปีก่อนหน้า แต่ถ้าพิจารณาแกนตั้งของแผนภูมิแท่ง จะเห็นว่าระยะไม่ได้เริ่มจาก 0 จึงอาจทำให้เข้าใจผิดว่ารายได้ของบริษัท A ทั้งสองปีแตกต่างกันมาก ทั้งที่จริงแล้วรายได้ของบริษัท A ใน พ.ศ. 2561 เพิ่มขึ้นจากปีก่อนหน้าเพียง 2.80 ล้านบาท หรือคิดเป็นร้อยละ 4.42 เท่านั้น ดังแสดงด้วยแผนภูมิแท่งต่อไปนี้



ในทำนองเดียวกัน การปรับระยะบนแกนนอนก็อาจก่อให้เกิดความเข้าใจคลาดเคลื่อนได้ เช่น การนำเสนอข้อมูลด้วยแผนภูมิเส้นดังต่อไปนี้



จะเห็นว่ากราฟมีลักษณะใกล้เคียงเส้นตรง อาจทำให้เข้าใจผิดว่ายอดขายของบริษัท A เพิ่มขึ้นในอัตราที่สม่ำเสมอตลอดปี แต่ถ้าปรับระยะห่างระหว่างเดือนต่าง ๆ บนแกนนอนให้สอดคล้องกับช่วงเวลาที่แท้จริง จะได้แผนภูมิเส้นดังรูป



จากกราฟ จะเห็นว่าที่จริงแล้วยอดขายของบริษัท A ไม่ได้เพิ่มขึ้นในอัตราที่สม่ำเสมอตลอดปี

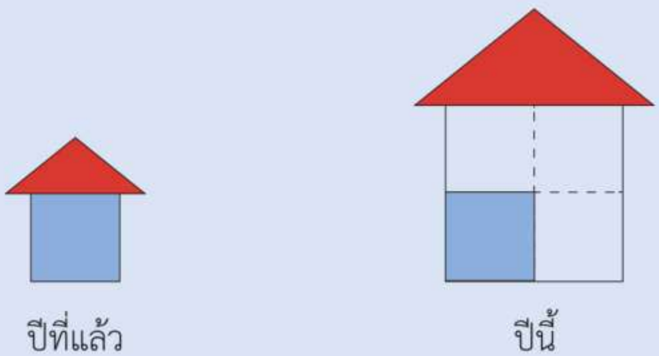
จากตัวอย่างทั้งหมดข้างต้น จะเห็นว่าความรู้ความเข้าใจเกี่ยวกับสถิติศาสตร์เบื้องต้นเป็นสิ่งจำเป็นอย่างยิ่ง ทั้งสำหรับผู้นำเสนอข้อมูล เพื่อให้สามารถนำเสนอภาพรวมที่ถูกต้อง เป็นประโยชน์ต่อสังคม และสำหรับผู้ใช้ข้อมูล เพื่อให้สามารถวิเคราะห์ข้อมูลข่าวสารที่พบตามสื่อต่าง ๆ ทั้งสื่อสิ่งพิมพ์และสื่ออิเล็กทรอนิกส์ได้ ตลอดจนรู้เท่าทันสื่อเหล่านั้นว่ามีความน่าเชื่อถือเพียงใด หรือมีจุดมุ่งหมายแอบแฝงหรือไม่

แบบฝึกหัด 1.1

1. จงยกตัวอย่างประโยชน์ของสถิติศาสตร์ที่เกี่ยวข้องกับชีวิตประจำวันของนักเรียนมาอย่างน้อย 3 ตัวอย่าง

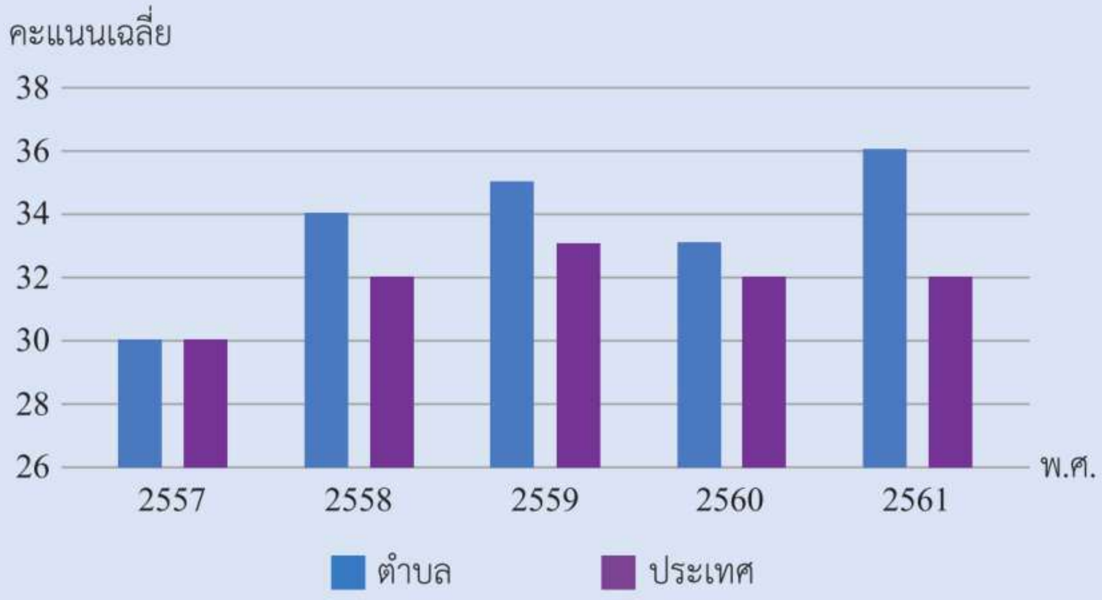
2. โรงเรียนแห่งหนึ่งเปิดสอนตั้งแต่ชั้นประถมศึกษาปีที่ 1 ถึงชั้นมัธยมศึกษาปีที่ 6 ถ้าโรงเรียนแห่งนี้ได้สอบถามความสูงของนักเรียนทั้งหมดในโรงเรียน พบว่าความสูงเฉลี่ยของนักเรียนทั้งหมด คือ 140 เซนติเมตร จึงสรุปว่านักเรียนส่วนใหญ่ของโรงเรียนแห่งนี้สูงประมาณ 140 เซนติเมตร จงพิจารณาว่าข้อสรุปดังกล่าวเหมาะสมหรือไม่ เพราะเหตุใด

3. ห้างสรรพสินค้าแห่งหนึ่งมีกำหนดเปิดให้บริการในอีก 4 เดือนข้างหน้า จึงได้ทำเอกสารเชิญชวนผู้สนใจให้เข้ามาเช่าพื้นที่ในห้างสรรพสินค้า โดยเนื้อหาตอนหนึ่งในเอกสารระบุว่า "ชุมชนบริเวณใกล้เคียงห้างสรรพสินค้ากำลังขยายตัวอย่างรวดเร็ว โดยคาดว่าจะมีจำนวนหลังคาเรือนในปีนี้เป็น 2 เท่าของปีที่แล้ว" เพื่อให้ผู้อ่านเข้าใจเนื้อหาดังกล่าวได้ดียิ่งขึ้น ในเอกสารจึงมีภาพประกอบ โดยทำรูปทางด้านขวาให้มีความกว้างและความยาวเป็น 2 เท่าของรูปทางด้านซ้าย เพื่อสื่อว่าจำนวนหลังคาเรือนในปีนี้จะเป็น 2 เท่าของปีที่แล้ว ดังนี้



จงพิจารณาว่าการนำเสนอข้อมูลด้วยภาพดังกล่าวก่อให้เกิดความเข้าใจคลาดเคลื่อนหรือไม่ เพราะเหตุใด

4. การสอบวัดผลวิชาคณิตศาสตร์ระดับประเทศที่จัดโดยสถาบันแห่งหนึ่ง มีคะแนนเต็ม 100 คะแนน โดยคะแนนสอบเฉลี่ยของนักเรียนในตำบลหนึ่งเทียบกับคะแนนสอบเฉลี่ยของนักเรียนทั่วประเทศ ใน พ.ศ. 2557 – 2561 แสดงได้ด้วยแผนภูมิแท่ง ดังนี้



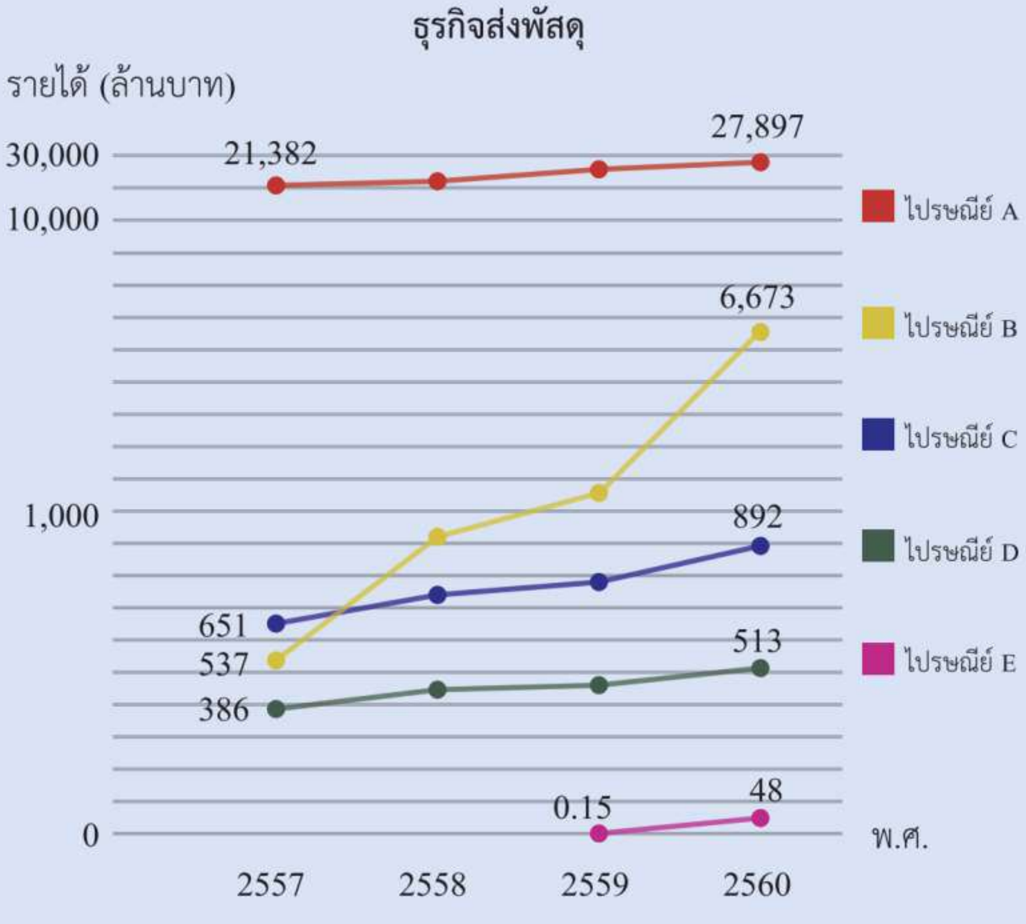
1) สามารถสรุปได้หรือไม่ว่าคะแนนสอบเฉลี่ยของนักเรียนในตำบลนี้ใน พ.ศ. 2558 เป็นสองเท่าของคะแนนสอบเฉลี่ยใน พ.ศ. 2557
2) จงสร้างแผนภูมิแท่งโดยให้ระยะบนแกนตั้งเริ่มจาก 0
3) เมื่อเปรียบเทียบแผนภูมิแท่งที่ได้ในข้อ 2) กับแผนภูมิแท่งที่โจทย์กำหนด จงพิจารณาว่าแผนภูมิแท่งที่โจทย์กำหนดก่อให้เกิดความเข้าใจคลาดเคลื่อนหรือไม่ เพราะเหตุใด

5. จงพิจารณาว่าการนำเสนอข้อมูลต่อไปนี้ก่อให้เกิดความเข้าใจคลาดเคลื่อนหรือไม่ เพราะเหตุใด

1)



2)



6. ให้นักเรียนค้นหาและเลือกข้อมูลข่าวสารที่สนใจจากสื่อสิ่งพิมพ์หรือสื่ออิเล็กทรอนิกส์ เช่น ผลสำรวจความคิดเห็น สถิติจำนวนประชากร มูลค่าการส่งออกสินค้า แล้วอธิบายเกี่ยวกับประโยชน์และความรู้ที่ได้รับจากข้อมูลดังกล่าว

7. ให้นักเรียนค้นหาและเลือกข้อมูลข่าวสารที่ก่อให้เกิดความเข้าใจคลาดเคลื่อน พร้อมทั้งอธิบายว่าข้อมูลข่าวสารดังกล่าวก่อให้เกิดความเข้าใจคลาดเคลื่อนอย่างไร

1.2 คำสำคัญในสถิติศาสตร์

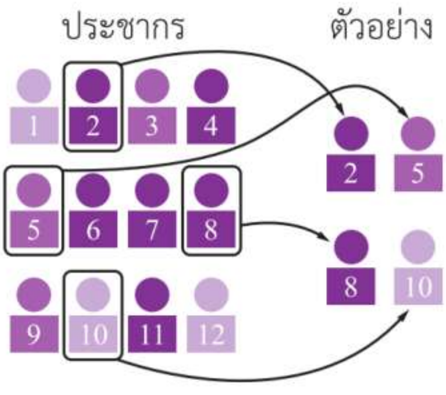
ประชากร (population) หมายถึง กลุ่มของหน่วยทั้งหมดในเรื่องที่สนใจศึกษา หน่วยในที่นี้อาจเป็น คน สัตว์ หรือสิ่งของ

ตัวอย่าง (sample) หมายถึง กลุ่มย่อยของประชากรที่ถูกเลือกมาเป็นตัวแทนของประชากร โดยทั่วไป มีวัตถุประสงค์เพื่อใช้ตัวอย่างในการสรุปผลเกี่ยวกับลักษณะของประชากรที่สนใจ

ตัวอย่างเช่น ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 อาจกำหนดตัวอย่างได้ดังนี้

1. ผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน
2. ผู้ป่วยโรคเบาหวานในภาคกลางของประเทศไทยใน พ.ศ. 2562

จากความหมายของประชากรและตัวอย่าง สามารถเขียนแสดงความสัมพันธ์ระหว่างประชากรและตัวอย่างได้ดังรูป



รูปที่ 1

ตัวแปร (variable) หมายถึง ลักษณะบางประการของประชากรหรือตัวอย่างที่สนใจศึกษา

ข้อมูล (data) หมายถึง ข้อความจริงเกี่ยวกับเรื่องใดเรื่องหนึ่งที่สามารถใช้ในการสรุปผลในเรื่องที่สนใจศึกษา อาจเป็นได้ทั้งตัวเลขหรือไม่ใช่ตัวเลข หรืออาจหมายถึงค่าของตัวแปรที่สนใจศึกษา

ตัวอย่างที่ 1

จากการเลือกตัวอย่างนักเรียนระดับมัธยมศึกษาตอนปลายของโรงเรียนแห่งหนึ่งจำนวน 50 คน เพื่อพิจารณาดัชนีมวลกาย (body mass index: BMI) ซึ่งใช้เป็นมาตรฐานในการประเมินสภาวะของร่างกายว่าอยู่ในเกณฑ์ที่เหมาะสมหรือไม่ โดยได้สำรวจข้อมูลต่าง ๆ ที่เกี่ยวข้อง ปรากฏผลดังตาราง

ลำดับที่	เลขประจำตัวนักเรียน	เพศ	อายุ (ปี)	น้ำหนัก (กิโลกรัม)	ส่วนสูง (เซนติเมตร)
1	45146	หญิง	17	70	170
2	48607	ชาย	16	45	155
3	50143	หญิง	18	50	168
⋮	⋮	⋮	⋮	⋮	⋮
50	46804	หญิง	16	55	159

จงระบุว่าประชากร ตัวอย่าง ตัวแปร และข้อมูลของการสำรวจนี้คืออะไร

วิธีทำ ประชากร คือ นักเรียนระดับมัธยมศึกษาตอนปลายทั้งหมดของโรงเรียนแห่งนี้

ตัวอย่าง คือ นักเรียนระดับมัธยมศึกษาตอนปลายของโรงเรียนแห่งนี้จำนวน 50 คน ที่เลือกมาเก็บข้อมูล

ตัวแปร คือ รหัสนักเรียน เพศ อายุ น้ำหนัก และส่วนสูง

ข้อมูล คือ ค่าของตัวแปรแต่ละตัวที่เก็บรวบรวมได้ ดังที่ปรากฏในบริเวณที่แรเงาในตาราง

■

พารามิเตอร์ (parameter) หมายถึง ค่าวัดที่แสดงลักษณะของประชากร ซึ่งเป็นค่าคงตัวที่คำนวณหรือประมวลจากข้อมูลทั้งหมดของประชากร

ตัวอย่างเช่น ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 พารามิเตอร์อาจเป็นได้ดังนี้

1. อายุเฉลี่ยของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562
2. ฐานนิยมของอาชีพของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562

ค่าสถิติ (statistic) เป็นค่าคงตัวที่พิจารณาจากข้อมูลของตัวอย่าง โดยมีวัตถุประสงค์เพื่ออธิบายลักษณะของตัวอย่างนั้นหรือเพื่อประมาณค่าของพารามิเตอร์แล้วนำไปใช้ในการอธิบายลักษณะของประชากร

ตัวอย่างเช่น ถ้ากำหนดให้ประชากรคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 และเลือกตัวอย่างคือผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน ค่าสถิติอาจเป็นได้ดังนี้

1. อายุเฉลี่ยของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน
2. ฐานนิยมของอาชีพของผู้ป่วยโรคเบาหวานในประเทศไทยใน พ.ศ. 2562 ที่สุ่มตัวอย่างมาจำนวน 10,000 คน

ตัวอย่างที่ 2

จากการสำรวจเกี่ยวกับอาชีพในฝันของนักเรียนในยุคดิจิทัลของจังหวัดหนึ่ง โดยสำรวจจากนักเรียนที่มีอายุ 15 – 18 ปี ที่เลือกมาจากทุกโรงเรียนในจังหวัด รวมทั้งสิ้น 300 คน พบว่า อาชีพในฝันของนักเรียนในยุคดิจิทัล 5 อันดับแรก ได้แก่ อันดับที่ 1 อาชีพธุรกิจส่วนตัว ร้อยละ 35 อันดับที่ 2 อาชีพครู ร้อยละ 22 อันดับที่ 3 อาชีพรับราชการ ร้อยละ 17 อันดับที่ 4 อาชีพแพทย์ ร้อยละ 12 และอันดับที่ 5 อาชีพวิศวกร ร้อยละ 7 จงระบุว่าประชากร ตัวอย่าง ตัวแปร ข้อมูล และค่าสถิติของการสำรวจนี้คืออะไร

วิธีทำ ประชากร คือ นักเรียนทุกคนในจังหวัดนี้

ตัวอย่าง คือ นักเรียนที่มีอายุ 15 – 18 ปี ที่เลือกมาจากทุกโรงเรียนในจังหวัดนี้ จำนวน 300 คน

ตัวแปร คือ อาชีพในฝันของนักเรียนในยุคดิจิทัล

ข้อมูล คือ อาชีพในฝันของนักเรียนในยุคดิจิทัลของนักเรียนแต่ละคนที่เก็บรวบรวมมาได้

ค่าสถิติ คือ ร้อยละของนักเรียนที่เลือกอาชีพในฝันในยุคดิจิทัล 5 อันดับแรก ซึ่งคำนวณจากตัวอย่างจำนวน 300 คน ■

ตัวอย่างที่ 3

จากการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน พ.ศ. 2560 สำนักงานสถิติแห่งชาติได้เก็บรวบรวมข้อมูลในช่วง 6 เดือนแรกของ พ.ศ. 2560 จากครัวเรือนตัวอย่างในทุกจังหวัดทั่วประเทศทั้งในเขตเทศบาลและนอกเขตเทศบาลประมาณ 26,000 ครัวเรือน เพื่อศึกษารายได้เฉลี่ยต่อเดือนของครัวเรือนทั่วประเทศ โดยจากข้อมูลตัวอย่างได้ข้อสรุปว่าค่าประมาณของรายได้เฉลี่ยต่อเดือนของครัวเรือนทั่วประเทศ คือ 26,973 บาท จงระบุว่าประชากร ตัวอย่าง พารามิเตอร์ ตัวแปร ข้อมูล และค่าสถิติของการสำรวจนี้คืออะไร

วิธีทำ ประชากร คือ ครัวเรือนทั้งหมดในประเทศไทยในช่วง 6 เดือนแรกของ พ.ศ. 2560

ตัวอย่าง คือ ครัวเรือนตัวอย่างในทุกจังหวัดทั่วประเทศทั้งในเขตเทศบาลและนอกเขตเทศบาลในช่วง 6 เดือนแรกของ พ.ศ. 2560 ที่เลือกมาเก็บข้อมูล ซึ่งมีจำนวนประมาณ 26,000 ครัวเรือน

ตัวแปร คือ รายได้เฉลี่ยต่อเดือนของครัวเรือน

ข้อมูล คือ รายได้เฉลี่ยต่อเดือนของแต่ละครัวเรือนตัวอย่างที่เก็บรวบรวมมาได้

พารามิเตอร์ คือ รายได้เฉลี่ยต่อเดือนของครัวเรือนทั่วประเทศ

ค่าสถิติ คือ ค่าประมาณของรายได้เฉลี่ยต่อเดือนของครัวเรือนทั่วประเทศ ซึ่งคือ 26,973 บาท เนื่องจากเป็นค่าที่ได้จากข้อมูลตัวอย่าง ■

แบบฝึกหัด 1.2

1. ในการศึกษาเกี่ยวกับความรู้เรื่องโรคเบาหวานและพฤติกรรมการดูแลตนเองของผู้ป่วยโรคเบาหวานของโรงพยาบาลแห่งหนึ่ง ผู้ศึกษาได้เก็บรวบรวมข้อมูลเกี่ยวกับเพศ อายุ น้ำหนัก ประวัติการเป็นโรคเบาหวานของคนในครอบครัว ความรู้เรื่องโรคเบาหวาน และพฤติกรรมการดูแลตนเอง โดยสุ่มตัวอย่างผู้ป่วยโรคเบาหวานจำนวน 120 คน จากโรงพยาบาลแห่งนี้ และใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูล
    1) จงระบุว่าประชากรและตัวอย่างของการศึกษานี้คืออะไร
    2) จงยกตัวอย่างตัวแปรของการศึกษานี้มาอย่างน้อย 3 ตัวแปร

2. สวนดุสิตโพล มหาวิทยาลัยสวนดุสิต ได้สำรวจความคิดเห็นของประชาชนที่พักอาศัยอยู่ในกรุงเทพมหานครและปริมณฑลจำนวนทั้งสิ้น 1,353 คน ระหว่างวันที่ 30 มกราคม – 1 กุมภาพันธ์ พ.ศ. 2562 เกี่ยวกับสถานการณ์ฝุ่นละออง ซึ่งในแบบสำรวจได้กำหนดตัวเลือก 5 ตัว และผู้ตอบแบบสำรวจสามารถเลือกตัวเลือกได้มากกว่า 1 ตัว ได้ผลสำรวจดังนี้

อันดับ 1	เข้าขั้นวิกฤต เกินมาตรฐาน ร้ายแรงมากขึ้นเรื่อย ๆ ยังแก้ไขไม่ได้	38.98%
อันดับ 2	คุณภาพชีวิตแย่ อันตรายต่อสุขภาพ ระบบทางเดินหายใจ	37.01%
อันดับ 3	ภาครัฐควรเร่งแก้ปัญหา มีมาตรการเร่งด่วน ทำอย่างจริงจัง	30.02%
อันดับ 4	ประชาชนต้องดูแลตัวเอง สวมหน้ากาก ไม่ทำกิจกรรมกลางแจ้ง	18.50%
อันดับ 5	ควบคุมยาก ทุกคนต้องช่วยกัน แก้ปัญหาระยะยาว	15.71%

จงระบุว่าตัวอย่าง ตัวแปร และข้อมูลของการสำรวจนี้คืออะไร

3. จากการสำรวจเกี่ยวกับเงินออมในแต่ละเดือนของนักเรียนระดับมัธยมศึกษาตอนปลายในจังหวัดหนึ่ง โดยสุ่มตัวอย่างนักเรียนระดับมัธยมศึกษาตอนปลายในจังหวัดนี้มาจำนวน 3,000 คน พบว่านักเรียนระดับมัธยมศึกษาตอนปลายในจังหวัดนี้มีเงินออมเฉลี่ยเดือนละ 700 บาท จงระบุว่าประชากร ตัวอย่าง ตัวแปร ข้อมูล พารามิเตอร์ และค่าสถิติของการสำรวจนี้คืออะไร

4. บริษัทผลิตตัวเก็บประจุไฟฟ้าแห่งหนึ่งต้องการตรวจสอบคุณภาพสินค้า โดยบริษัทได้ผลิตตัวเก็บประจุไฟฟ้าล็อตละ 500 ชิ้น แล้วสุ่มตัวอย่างตัวเก็บประจุไฟฟ้าจำนวน 50 ชิ้นจากแต่ละล็อตมาตรวจสอบคุณภาพสินค้า ถ้าพบว่ามีตัวเก็บประจุไฟฟ้าชำรุดอย่างน้อย 1 ชิ้นจะไม่จำหน่ายสินค้าล็อตนั้น จงระบุว่าประชากร ตัวอย่าง ตัวแปร และข้อมูลของการสำรวจในแต่ละล็อตคืออะไร

1.3 ประเภทของข้อมูล

ข้อมูลที่จะนำมาใช้ศึกษาสามารถแบ่งได้หลายประเภทที่สำคัญมีดังนี้

1. การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล
2. การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บ
3. การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูล

1.3.1 การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูล

การแบ่งประเภทของข้อมูลตามแหล่งที่มาของข้อมูลเป็นการแบ่งประเภทของข้อมูล โดยคำนึงว่า ผู้ใช้ข้อมูลเป็นผู้จัดเก็บข้อมูลเอง หรือเป็นข้อมูลที่บุคคลหรือหน่วยงานอื่นเป็นผู้จัดเก็บแล้วผู้ใช้เพียงแค่นำมาใช้ จึงแบ่งข้อมูลออกเป็น 2 ประเภท ดังนี้

ข้อมูลปฐมภูมิ (primary data) คือข้อมูลที่ผู้ใช้ดำเนินการเก็บรวบรวมจากเจ้าของข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง

ตัวอย่างของข้อมูลปฐมภูมิ

1. ข้อมูลความพึงพอใจในสินค้าที่เจ้าของสินค้าเก็บรวบรวมจากผู้ใช้สินค้าหรือผู้บริโภค เป็นข้อมูลปฐมภูมิของเจ้าของสินค้า
2. ข้อมูลการรักษาพยาบาลที่โรงพยาบาลบันทึกไว้ในประวัติผู้ป่วย เป็นข้อมูลปฐมภูมิของโรงพยาบาล
3. ข้อมูลที่นักเรียนบันทึกจากการทดลองฟิสิกส์ในห้องปฏิบัติการของโรงเรียน เป็นข้อมูลปฐมภูมิของนักเรียนที่ทำการทดลอง

ข้อมูลทุติยภูมิ (secondary data) คือข้อมูลที่ผู้ใช้ไม่ได้ดำเนินการเก็บรวบรวมจากเจ้าของข้อมูลหรือต้นกำเนิดของข้อมูลโดยตรง แต่ใช้ข้อมูลที่บุคคลหรือหน่วยงานอื่นเก็บรวบรวมมา ซึ่งส่วนใหญ่ผู้ใช้มักจะใช้ข้อมูลที่เก็บรวบรวมโดยภาครัฐซึ่งเป็นการเก็บรวบรวมข้อมูลตามภารกิจของหน่วยงาน

ตัวอย่างของข้อมูลทุติยภูมิ

1. สำนักงานตรวจคนเข้าเมืองจัดเก็บข้อมูลนักท่องเที่ยวที่เดินทางเข้ามาในประเทศไทยในช่วงไตรมาสหนึ่ง ข้อมูลนี้เป็นข้อมูลปฐมภูมิของสำนักงานตรวจคนเข้าเมือง แต่เป็นข้อมูลทุติยภูมิสำหรับผู้วิจัยหรือผู้ศึกษาอื่นที่นำข้อมูลนี้มาวิเคราะห์เพื่อหาคำตอบที่สนใจ
2. โรงเรียนจัดเก็บข้อมูลทะเบียนประวัติของนักเรียน ข้อมูลนี้เป็นข้อมูลปฐมภูมิของโรงเรียนเนื่องจากโรงเรียนเป็นผู้จัดเก็บจากนักเรียนซึ่งเป็นเจ้าของข้อมูล แต่เมื่อโรงเรียนส่งรายงานข้อมูลไปยังเขตพื้นที่การศึกษา ข้อมูลนี้จะเป็นข้อมูลทุติยภูมิของเขตพื้นที่การศึกษา

ข้อดีของข้อมูลทุติยภูมิคือผู้ใช้ไม่ต้องเสียเวลาและค่าใช้จ่ายในการเก็บรวบรวมข้อมูลเองสามารถนำข้อมูลที่มีผู้อื่นเก็บรวบรวมไว้แล้วมาใช้ได้เลย แต่อย่างไรก็ตามผู้ใช้จะต้องระมัดระวังในการนำข้อมูลประเภทนี้มาใช้ เนื่องจากมีโอกาสผิดพลาดได้ง่าย ดังนั้น การนำข้อมูลทุติยภูมิมาใช้จึงมีสิ่งสำคัญที่ควรจะพิจารณาดังต่อไปนี้

1. บุคคลหรือหน่วยงานที่จัดทำรายงาน บทความ หรือเอกสาร มีความรู้ความเชี่ยวชาญในเรื่องนั้น และมีความน่าเชื่อถือหรือไม่ การเขียนอาศัยเหตุผลและหลักวิชาการมากน้อยเพียงใด และข้อมูลที่นำมาใช้ควรเป็นข้อมูลที่ผู้เขียนเก็บรวบรวมมาเองโดยตรง เช่น ข้อมูลที่ได้จากการสำรวจหรือการสำมะโน หากไม่มีความจำเป็นไม่ควรใช้ข้อมูลที่ผู้เขียนนำมาจากแหล่งข้อมูลอื่น เนื่องจากอาจมีการคลาดเคลื่อนจากความเป็นจริงได้มาก
2. ควรเก็บรวบรวมข้อมูลจากหลาย ๆ แหล่ง ถ้าสามารถทำได้ เพื่อใช้ในการตรวจสอบว่าข้อมูลที่ต้องการมีความผิดพลาดจากการคัดลอกหรือเข้าใจผิดหรือไม่ นอกจากนี้ ผู้เก็บรวบรวมข้อมูลควรใช้ความรู้ความชำนาญของตนเองในเรื่องนั้นเพื่อพิจารณาว่าข้อมูลที่ได้มามีความใกล้เคียงกับความเป็นจริงหรือไม่
3. พิจารณาจากลักษณะของข้อมูล ถ้าเป็นข้อมูลที่เป็นข้อความจริง ข้อมูลที่ได้จากทะเบียนหรือข้อมูลที่เป็นความคิดเห็นหรือเจตคติ ส่วนใหญ่มักจะมีความถูกต้องและเชื่อถือได้มาก แต่ถ้าเป็นข้อมูลที่เป็นเรื่องส่วนตัวหรือข้อมูลซึ่งผู้ตอบอาจต้องเสียประโยชน์จากการตอบข้อมูลอาจจะไม่ถูกต้องและเชื่อถือได้น้อย
4. ถ้าข้อมูลที่เก็บรวบรวมได้มาจากการสำรวจตัวอย่าง ควรจะต้องตรวจสอบวิธีการที่ใช้ในการเลือกตัวอย่าง ขนาดตัวอย่าง และวิธีการวิเคราะห์ข้อมูลว่าเหมาะสมหรือไม่

1.3.2 การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บ

การแบ่งประเภทของข้อมูลตามระยะเวลาที่จัดเก็บเป็นการแบ่งประเภทของข้อมูลโดยพิจารณาจากช่วงเวลาที่มีข้อมูลเกิดขึ้นและมีการจัดเก็บ โดยแบ่งประเภทของข้อมูลออกเป็นข้อมูลอนุกรมเวลาและข้อมูลตัดขวาง

ข้อมูลอนุกรมเวลา (time series data) คือชุดข้อมูลที่เกิดขึ้นและจัดเก็บตามลำดับเวลาต่อเนื่องกันไปตลอดช่วง ๆ หนึ่ง

ข้อมูลอนุกรมเวลาสามารถแสดงการเปลี่ยนแปลงหรือการเคลื่อนไหวของข้อมูลที่สนใจเมื่อเวลาเปลี่ยนไป ข้อมูลประเภทนี้จึงเป็นประโยชน์สำหรับการทำวิจัยระยะยาว เนื่องจากทำให้ผู้วิจัยเห็นแนวโน้มของเรื่องนั้นได้ ตัวอย่างเช่น ยอดขายสินค้ารายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2558 ถึงเดือนธันวาคม พ.ศ. 2561 คือยอดขายสินค้าที่เกิดขึ้นและมีการบันทึกในแต่ละเดือนว่าเป็นเท่าใดตลอดช่วงเวลานั้น ซึ่งเมื่อนำข้อมูลรายเดือนมาเรียงต่อกันตามลำดับเวลา จะแสดงการเคลื่อนไหวขึ้นลงของยอดขายสินค้ารายเดือนในช่วงนั้นว่าเป็นอย่างไร เดือนใดมียอดขายสูงสุดและเดือนใดมียอดขายต่ำสุด ลักษณะการเปลี่ยนแปลงหรือเคลื่อนไหวเมื่อเวลาเปลี่ยนแปลงไปเป็นอย่างไร

ข้อมูลตัดขวาง (cross-sectional data) คือข้อมูลที่บอกสถานะหรือสภาพของสิ่งที่สนใจ ณ จุดหนึ่งของเวลา

ตัวอย่างของข้อมูลตัดขวาง

1. จำนวนประชากรของประเทศไทย ณ วันที่ 13 เมษายน พ.ศ. 2562 โดยอาจแสดงว่ามีประชากรรวมทั้งสิ้นกี่คน เป็นเพศชายและเพศหญิงกี่คน เป็นประชากรอายุน้อยกว่า 1 ปี กี่คน เป็นประชากรอายุ 1 – 5 ปี กี่คน เป็นประชากรสูงอายุคือตั้งแต่ 65 ปีขึ้นไปกี่คน อาศัยอยู่ในแต่ละจังหวัดจำนวนเท่าใด
2. รายงานผลการศึกษาของนักเรียนเมื่อสิ้นภาคการศึกษา เนื่องจากเป็นการแสดงสถานภาพทางการเรียนของนักเรียน ณ วันประกาศผลเมื่อสิ้นภาคการศึกษา

1.3.3 การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูล

การแบ่งประเภทของข้อมูลตามลักษณะของข้อมูลเป็นการแบ่งประเภทของข้อมูลโดยพิจารณาว่าข้อมูลนั้นแสดงถึงปริมาณของสิ่ง ๆ หนึ่งหรือไม่ โดยแบ่งออกได้เป็นข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพ

ข้อมูลเชิงปริมาณ (quantitative data) คือข้อมูลที่ได้จากการวัดหรือการนับค่า โดยแสดงเป็นตัวเลขหรือปริมาณที่สามารถนำไปบวก ลบ คูณ หรือหาร และเปรียบเทียบกันได้

ตัวอย่างของข้อมูลเชิงปริมาณ เช่น คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องหนึ่ง จำนวนนักเรียนที่ใช้บริการห้องสมุดของโรงเรียนในแต่ละวันในภาคการศึกษาที่แล้ว ยอดขายรถยนต์รายเดือน

ข้อมูลเชิงคุณภาพ (qualitative data) คือข้อมูลที่แสดงลักษณะ ประเภท สมบัติในเชิงคุณภาพ และอื่น ๆ ที่ไม่สามารถวัดค่าเป็นตัวเลขที่นำมาบวก ลบ คูณ หรือหารกันได้

ตัวอย่างของข้อมูลเชิงคุณภาพ เช่น อาชีพของผู้ปกครองนักเรียนห้องหนึ่ง หมายเลขโทรศัพท์ของนักเรียนห้องหนึ่ง ความคิดเห็นที่ประชาชนในท้องถิ่นมีต่อเรื่องหนึ่ง ๆ ความพึงพอใจในการใช้บริการห้องสมุดของนักเรียนว่าพอใจมากน้อยเพียงไร (พอใจมากที่สุด พอใจมาก พอใจปานกลาง พอใจน้อย พอใจน้อยที่สุด)

ข้อมูลเชิงคุณภาพโดยทั่วไปจะใช้วิธีการวิเคราะห์ข้อมูลที่แตกต่างไปจากข้อมูลเชิงปริมาณ อย่างไรก็ตาม สำหรับข้อมูลเชิงปริมาณ อาจจัดกลุ่มหรือช่วงของค่าเพื่อให้ตีความเป็นข้อมูลเชิงคุณภาพได้ ตัวอย่างเช่น ผลผลิตข้าวต่อไร่ในพื้นที่บริเวณหนึ่งมีค่าที่เป็นไปได้ตั้งแต่ 295 ถึง 560 กิโลกรัม โดยมีค่าเฉลี่ยอยู่ที่ 400 กิโลกรัม อาจเปลี่ยนข้อมูลปริมาณผลผลิตข้าวต่อไร่เป็นข้อมูลเชิงคุณภาพที่บอกเพียงระดับผลผลิตว่าต่ำ ปานกลาง หรือสูงได้ โดยกำหนดให้ผลผลิตต่ำคือผลผลิตตั้งแต่ 295 กิโลกรัม แต่น้อยกว่า 375 กิโลกรัม ผลผลิตปานกลางคือผลผลิตตั้งแต่ 375 กิโลกรัม แต่น้อยกว่า 450 กิโลกรัม และผลผลิตสูงคือผลผลิตตั้งแต่ 450 ถึง 560 กิโลกรัม แต่ข้อมูลที่บันทึกใหม่จะไม่มีรายละเอียดของข้อมูลมากเท่ากับข้อมูลเดิมและต้องใช้วิธีการวิเคราะห์ที่แตกต่างออกไปด้วย

ข้อมูลที่นำมาประมวลผลหรือวิเคราะห์เพื่อตอบคำถามหรือโจทย์ที่ตั้งไว้ อาจแบ่งตามประเภทต่าง ๆ ข้างต้นได้มากกว่าหนึ่งประเภท เช่น ข้อมูลหนึ่งอาจเป็นทั้งข้อมูลเชิงปริมาณและข้อมูลอนุกรมเวลา เช่น จำนวนนักท่องเที่ยวชาวต่างประเทศที่เดินทางเข้ามาที่สนามบินสุวรรณภูมิในแต่ละวันในเดือนมกราคม พ.ศ. 2560

แบบฝึกหัด 1.3

1. ในการวิจัยเกี่ยวกับพฤติกรรมและปัจจัยทำนายพฤติกรรมการป้องกันโรคความดันเลือดสูงในกลุ่มประชาชนที่มีภาวะก่อนเป็นโรคความดันเลือดสูง ตัวอย่างที่ใช้ศึกษาคือประชาชนจำนวน 250 คน ที่มีค่าความดันเลือดอยู่ระหว่าง 120 – 139 / 80 – 89 มิลลิเมตรปรอท เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลประกอบด้วยแบบสอบถามข้อมูลส่วนบุคคล แบบสอบถามความรู้เกี่ยวกับโรคความดันเลือดสูง และแบบสอบถามเจตคติต่อโรคและพฤติกรรมการป้องกันโรคความดันเลือดสูง จงพิจารณาว่าข้อมูลที่ได้จากการวิจัยข้างต้นเป็นข้อมูลปฐมภูมิหรือข้อมูลทุติยภูมิ

2. โรงงานผลิตอาหารสำเร็จรูปแห่งหนึ่งใช้ข้อมูลจากสำนักงานสถิติแห่งชาติดังตารางด้านล่าง ในการวางแผนการผลิตอาหารว่าจะปรุงด้วยวิธีใด เพื่อนำไปจำหน่ายในร้านสะดวกซื้อ

วิธีปรุงอาหารที่รับประทานเป็นประจำ	กลุ่มอายุ				
	รวม	6 - 14 ปี	15 - 24 ปี	25 - 59 ปี	60 ปี ขึ้นไป
ต้ม/ลวกสุก	50.4	32.6	37.3	51.1	71.1
ตุ๋น/นึ่ง/อบ	4.3	2.5	2.7	4.4	6.5
ผัด	27.4	25.5	34.8	29.8	14.7
ทอด	12.5	36.2	19.1	8.6	3.9
ลวกแบบสุก ๆ ดิบ ๆ	0.6	0.1	0.5	0.8	0.2
ปิ้ง/ย่าง	4.1	2.9	5.0	4.5	2.9
อื่น ๆ	0.7	0.2	0.6	0.8	0.7

จากสถานการณ์ข้างต้น จงพิจารณาว่าข้อมูลที่โรงงานผลิตอาหารสำเร็จรูปนำมาใช้เป็นข้อมูลปฐมภูมิหรือข้อมูลทุติยภูมิ

3. จงพิจารณาว่าหากต้องการเก็บข้อมูลต่อไปนี้ ควรใช้ข้อมูลอนุกรมเวลาหรือข้อมูลตัดขวาง
   1) ว่าที่นายกรัฐมนตรีในดวงใจของคนไทย พ.ศ. 2566
   2) ความหนาแน่นของประชากรต่อพื้นที่ในภาคเหนือ พ.ศ. 2562
   3) ราคาทองคำในตลาดโลกระหว่าง พ.ศ. 2550 – 2562
   4) ความสูงของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง เมื่อเปิดภาคเรียนที่ 1
   5) คณะยอดนิยมที่เด็กไทยเลือกเข้าศึกษาต่อในระดับอุดมศึกษาในช่วง 3 ปีที่ผ่านมา

4. จงพิจารณาว่าข้อมูลต่อไปนี้เป็นข้อมูลเชิงปริมาณหรือข้อมูลเชิงคุณภาพ
   1) คะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องหนึ่งโดยมีการแบ่งระดับคะแนนเป็น สูง กลาง และต่ำ
   2) จำนวนผู้โดยสารที่ใช้รถประจำทางในแต่ละวัน
   3) หมายเลขทะเบียนรถยนต์นั่งส่วนบุคคล
   4) รหัสไปรษณีย์ในกรุงเทพมหานคร
   5) ราคาข้าวสารต่อกิโลกรัม
   6) เลขประจำตัวประชาชนของพนักงานในบริษัทแห่งหนึ่ง
   7) บ้านเลขที่ของนักเรียนในโรงเรียนแห่งหนึ่ง
   8) รายได้ของสมาชิกในครอบครัวหนึ่ง
   9) สถานะการป่วยเป็นโรคหัวใจของคนไทย
   10) ความพึงพอใจของผู้บริโภคที่มีต่อร้านค้าแห่งหนึ่ง

5. ให้นักเรียนค้นหาและเลือกข้อมูลที่สนใจมา 5 รายการ พร้อมทั้งระบุประเภทของข้อมูล (ข้อมูลอนุกรมเวลา/ข้อมูลตัดขวาง และข้อมูลเชิงปริมาณ/ข้อมูลเชิงคุณภาพ)

1.4 สถิติศาสตร์เชิงพรรณนาและสถิติศาสตร์เชิงอนุมาน

การวิเคราะห์ข้อมูลแบ่งได้เป็นสถิติศาสตร์เชิงพรรณนาและสถิติศาสตร์เชิงอนุมาน โดยมีรายละเอียดดังต่อไปนี้

สถิติศาสตร์เชิงพรรณนา (descriptive statistics) เป็นการวิเคราะห์ข้อมูลที่สรุปสาระสำคัญของข้อมูลชุดหนึ่ง ซึ่งเป็นข้อมูลเชิงคุณภาพหรือข้อมูลเชิงปริมาณ เพื่ออธิบายลักษณะหรือสภาพของข้อมูลชุดนั้นว่าเป็นอย่างไร โดยทั่วไปข้อมูลเชิงคุณภาพจะใช้การนำเสนอด้วยตารางความถี่ แผนภูมิแท่ง ฐานนิยม และอื่น ๆ ส่วนข้อมูลเชิงปริมาณจะใช้การนำเสนอด้วยฮิสโทแกรม แผนภาพกล่อง ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย และอื่น ๆ

ตัวอย่างของการใช้สถิติศาสตร์เชิงพรรณนา

1. คะแนนต่ำสุดในการสอบกลางภาควิชาคณิตศาสตร์ คือ 7 คะแนน
2. นักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งในกรุงเทพมหานครที่มีภูมิลำเนาอยู่ทางภาคเหนือคิดเป็นร้อยละ 15 ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ทั้งหมดของโรงเรียนแห่งนี้

โดยทั่วไปข้อมูลที่เก็บรวบรวมมาได้อาจมีจำนวนมากหรือเก็บรวบรวมไม่เป็นระบบจึงไม่สามารถนำมาอธิบายลักษณะที่สำคัญของข้อมูลได้ชัดเจน จึงต้องใช้วิธีการของสถิติศาสตร์เชิงพรรณนาในการสรุปผลจากข้อมูลด้วยการนำเสนอข้อมูลในรูปแบบต่าง ๆ เพื่อให้สามารถอ่านและแปลความหมายของข้อมูลได้ง่ายขึ้น เช่น คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 50 คน เก็บรวบรวมมาได้ดังนี้

81	79	74	50	47	53	66	62	98	70
77	73	86	73	52	69	85	64	54	78
81	80	74	95	70	72	93	62	58	91
50	69	45	85	82	78	68	78	67	49
49	55	67	67	89	58	53	55	90	59

จากข้อมูลข้างต้น อาจใช้วิธีของสถิติศาสตร์เชิงพรรณนาในการนำเสนอข้อมูลด้วยตาราง โดยแบ่งตามช่วงคะแนนได้ดังนี้

คะแนนสอบ	จำนวนนักเรียน (คน)
41 – 50	6
51 – 60	9
61 – 70	12
71 – 80	11
81 – 90	8
91 – 100	4
รวม	**50**

จะเห็นว่าการนำเสนอข้อมูลด้วยตารางจะทำให้อ่านและแปลความหมายของข้อมูลได้ง่ายกว่าการพิจารณาจากข้อมูลที่เก็บรวบรวมมาได้ทั้งหมดโดยที่ยังไม่ได้จัดกลุ่มข้อมูลในรูปของตาราง

การเก็บรวบรวมข้อมูลโดยทั่วไปอาจมีข้อจำกัดทางด้านเวลาและทรัพยากรที่มี ทำให้ไม่สามารถเก็บรวบรวมข้อมูลที่สนใจจากทุกหน่วยในขอบข่ายที่ศึกษาได้ทั้งหมด ข้อมูลที่นำมาใช้จึงเป็นเพียงส่วนหนึ่งของประชากรเท่านั้น เช่น ปริมาณน้ำตาลในชานมไข่มุก 1 แก้ว จากแต่ละร้านในกรุงเทพมหานคร หากต้องการเก็บรวบรวมข้อมูลปริมาณน้ำตาลในชานมไข่มุก 1 แก้ว จากทุกร้านในกรุงเทพมหานคร อาจต้องใช้เวลา กำลังคน และงบประมาณจำนวนมาก ดังนั้น อาจจะเก็บรวบรวมข้อมูลปริมาณน้ำตาลในชานมไข่มุก 1 แก้ว จากบางร้านในกรุงเทพมหานคร แล้วใช้สถิติศาสตร์เชิงอนุมานในการหาข้อสรุปเกี่ยวกับปริมาณน้ำตาลในชานมไข่มุก 1 แก้ว จากทุกร้านในกรุงเทพมหานครได้

สถิติศาสตร์เชิงอนุมาน (inferential statistics) เป็นการวิเคราะห์ข้อมูลที่ใช้ทฤษฎีที่เกี่ยวกับความน่าจะเป็นในการหาข้อสรุปเกี่ยวกับลักษณะของประชากรโดยใช้ข้อมูลจากตัวอย่างที่ได้มาจากประชากรนั้น

ตัวอย่างของการใช้สถิติศาสตร์เชิงอนุมาน

ในการสำรวจพฤติกรรมการเดินทางท่องเที่ยวของคนไทยที่มีอายุ 15 ปีขึ้นไป ใน พ.ศ. 2559 ซึ่งจัดทำโดยสำนักงานสถิติแห่งชาติร่วมกับการท่องเที่ยวแห่งประเทศไทย (ททท.) ได้มีการเก็บรวบรวมข้อมูลในเดือนมกราคม – กุมภาพันธ์ พ.ศ. 2560 จากคนไทยที่มีอายุ 15 ปีขึ้นไปทั่วประเทศ ที่เลือกเป็นตัวอย่างจำนวน 63,060 คน และได้ข้อสรุปดังแสดงในรูปที่ 2 ซึ่งอาจสรุปได้ว่าวัตถุประสงค์หลักอันดับที่ 1 ในการเดินทางท่องเที่ยวของคนไทยทั้งประเทศที่มีอายุ 15 ปีขึ้นไป ใน พ.ศ. 2559 คือการเยี่ยมครอบครัว/ญาติ/เพื่อน ซึ่งคิดเป็น 33.1%

ร้อยละของคนไทยที่มีอายุ 15 ปีขึ้นไป ที่เดินทางท่องเที่ยวจำแนกตามวัตถุประสงค์หลัก
ในการเดินทางท่องเที่ยว 5 อันดับแรก ในรอบปี 2559



1. เยี่ยมครอบครัว/ญาติ/เพื่อน 33.1%
2. ท่องเที่ยว/พักผ่อน 20.2%
3. ไหว้พระ/ปฏิบัติธรรม 11.7%
4. ตั้งใจหรือแวะไปรับประทานอาหาร 9.0%
5. ซื้อของ/ชอปปิง 8.9%

ที่มาของภาพที่ 1 – 3, 5 : Freepik.com

รูปที่ 2

จากที่กล่าวมา จะเห็นว่าสถิติศาสตร์ครอบคลุมเรื่องของข้อมูลและการจัดการกับข้อมูลเพื่อให้ได้ผลสรุปที่สามารถนำไปใช้ประโยชน์ได้ โดยมีองค์ประกอบที่สำคัญ เช่น การเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล และการสรุปผลจากข้อมูลที่เกี่ยวข้อง

แบบฝึกหัด 1.4

จงพิจารณาว่าตัวอย่างต่อไปนี้ใช้วิธีการของสถิติศาสตร์เชิงพรรณนาหรือสถิติศาสตร์เชิงอนุมาน พร้อมทั้งให้เหตุผลประกอบ

1. สำนักโพลแห่งหนึ่งได้สำรวจเกี่ยวกับอาชีพในฝันของเยาวชนไทยใน พ.ศ. 2562 โดยสำรวจจากเยาวชนไทยที่มีอายุ 12 ปีขึ้นไป กระจายทุกภูมิภาค ระดับการศึกษา และอาชีพ รวมทั้งสิ้น 15,000 คน ซึ่งสรุปผลได้ว่าอาชีพในฝันของเยาวชนไทยทั้งประเทศ 5 อันดับแรก ได้แก่ อันดับที่ 1 อาชีพธุรกิจส่วนตัว ร้อยละ 16.40 อันดับที่ 2 อาชีพครู ร้อยละ 13.65 อันดับที่ 3 อาชีพรับราชการ ร้อยละ 13.36 อันดับที่ 4 อาชีพแพทย์ ร้อยละ 10.80 อันดับที่ 5 อาชีพวิศวกร ร้อยละ 7.53

2. จากการจัดทำสำมะโนประชากรและเคหะของสำนักงานสถิติแห่งชาติ พบว่า ใน พ.ศ. 2553 จังหวัดที่มีความหนาแน่นของประชากรโดยเฉลี่ยสูงสุด คือ กรุงเทพมหานคร รองลงมา คือ จังหวัดนนทบุรี สมุทรปราการ สมุทรสาคร ภูเก็ต ปทุมธานี สมุทรสงคราม นครปฐม ชลบุรี และพระนครศรีอยุธยา ตามลำดับ

เสริมสมอง : การสำมะโน

การสำมะโน (census) คือการเก็บรวบรวมข้อมูลเกี่ยวกับประชากร เคหะ การเกษตร อุตสาหกรรม ธุรกิจ และการอื่น ๆ เพื่อใช้ประโยชน์ในทางสถิติ โดยการแจงนับจากทุกหน่วยเกี่ยวกับเรื่องนั้น ๆ

3. ห้างสรรพสินค้าแห่งหนึ่งได้สำรวจความพึงพอใจของลูกค้าในด้านกิจกรรมส่งเสริมการตลาด เพื่อเป็นแนวทางในการวางแผนกลยุทธ์ทางการตลาด โดยสำรวจจากลูกค้าที่มาใช้บริการที่ห้างสรรพสินค้าในเดือนมกราคม พ.ศ. 2562 ที่เลือกเป็นตัวอย่างจำนวน 142 คน สรุปผลได้ว่า ร้อยละ 50 ของลูกค้าทุกคนที่มาใช้บริการห้างสรรพสินค้าแห่งนี้มีความพึงพอใจในระดับมากที่สุด

4. จากรายงานผลการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) วิชาภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2560 สามารถแสดงคะแนนเฉลี่ยรายภูมิภาค (จากคะแนนเต็ม 100 คะแนน) ได้ดังตาราง

ภูมิภาค	คะแนนเฉลี่ย
ภาคตะวันออกเฉียงเหนือ	24.36
ภาคใต้	26.49
ภาคตะวันตก	26.84
ภาคเหนือ	28.84
ภาคกลาง	28.85
ภาคตะวันออก	29.46
กรุงเทพมหานคร	39.65

5. สำนักโพลแห่งหนึ่งได้สำรวจเกี่ยวกับนโยบายที่สำคัญและเร่งด่วนที่สุดที่ประชาชนไทยต้องการจากรัฐบาลชุดใหม่ โดยสำรวจจากประชาชนไทยที่มีอายุ 18 ปีขึ้นไป กระจายทุกภูมิภาค ระดับการศึกษา และอาชีพ รวมทั้งสิ้น 10,000 คน ซึ่งสรุปผลได้ว่า ประชาชนส่วนใหญ่ของทั้งประเทศต้องการให้รัฐบาลชุดใหม่แก้ปัญหาปากท้องและหนี้สินของประชาชน ร้อยละ 55.75 รองลงมาต้องการให้ส่งเสริมราคาสินค้าและพืชผลทางการเกษตร ร้อยละ 28.23 ส่งเสริมโครงการเรียนฟรีถึงระดับปริญญาตรี ร้อยละ 5.76 แก้ไขปัญหาการทุจริตคอร์รัปชัน ร้อยละ 3.54 แก้ปัญหาหนี้นอกระบบ ร้อยละ 1.75 แก้ไขปัญหาด้านการคมนาคม ร้อยละ 1.68 และอื่น ๆ ร้อยละ 3.29

6. สำนักอนามัยการเจริญพันธุ์ กรมอนามัย กระทรวงสาธารณสุข ได้รายงานอัตราการคลอดบุตรในวัยรุ่นอายุ 10 – 14 ปี ต่อประชากรหญิงอายุ 10 – 14 ปี จำนวน 1,000 คน ดังตาราง

พ.ศ.	หญิงคลอดอายุ 10 – 14 ปี (คน)	หญิงอายุ 10 – 14 ปี ทั้งหมด (คน)	อัตรา : 1,000
2546	1,791	2,388,624	0.7
2547	2,469	2,382,623	1.0
2548	2,586	2,369,156	1.1
2549	2,545	2,384,031	1.1
2550	2,654	2,395,156	1.1
2551	2,745	2,384,707	1.2
2552	2,928	2,339,177	1.3
2553	3,074	2,272,507	1.4
2554	3,417	2,196,350	1.6
2555	3,710	2,096,028	1.8
2556	3,415	2,024,332	1.7
2557	3,213	1,991,041	1.6
2558	2,988	1,963,728	1.5
2559	2,746	1,941,436	1.4

7. สำนักงานสถิติแห่งชาติได้จัดทำโครงการ “สำรวจพฤติกรรมการออมและการเข้าถึงบริการทางการเงินภาคครัวเรือน” โดยได้ดำเนินการสำรวจไปพร้อมกับการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือนจากครัวเรือนตัวอย่างในทุกจังหวัดทั่วประเทศทั้งในเขตเทศบาลและนอกเขตเทศบาล จำนวน 13,000 ครัวเรือน โดยเก็บรวบรวมข้อมูลระหว่างเดือนกรกฎาคม – กันยายน พ.ศ. 2561 ซึ่งสรุปผลได้ว่า ครัวเรือนทั่วประเทศส่วนใหญ่มีวัตถุประสงค์หลักของการเก็บออมเพื่อใช้จ่ายในยามชราหรือเกษียณอายุ ร้อยละ 42.1 เพื่อใช้จ่ายยามเจ็บป่วย/ฉุกเฉิน ร้อยละ 32.0 เพื่อใช้จ่ายเมื่อเว้นว่างจากฤดูกาลเก็บเกี่ยว/ทำการเกษตร/ทำธุรกิจ ร้อยละ 9.7 เพื่อใช้ในการศึกษา ร้อยละ 7.9 เพื่อจัดหาเครื่องอำนวยความสะดวกและยานพาหนะ ร้อยละ 4.2 เพื่อมีบ้านเป็นของตัวเอง ร้อยละ 3.7 และอื่น ๆ ร้อยละ 0.4 เช่น ค้ำประกันเงินกู้ มรดก ท่องเที่ยว

8. โรงงานผลิตผลไม้กระป๋องผลิตเงาะกระป๋องล็อตละ 5,000 กระป๋อง ถ้าโรงงานเลือกตัวอย่างเงาะกระป๋องมาตรวจสอบจำนวน 100 กระป๋อง แล้วพบว่าเงาะกระป๋องที่เลือกมาได้มาตรฐานตามข้อกำหนด โรงงานแห่งนี้จึงสรุปว่าเงาะกระป๋องทุกกระป๋องในล็อตนี้ได้มาตรฐาน

กิจกรรม : แบบสอบถามพื้นฐานทางสังคมและเศรษฐกิจของนักเรียน

โรงเรียนแห่งหนึ่งต้องการเก็บข้อมูลพื้นฐานทางสังคมและเศรษฐกิจของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ทั้งหมดในโรงเรียน โดยให้ผู้ปกครองของนักเรียนทุกคนตอบแบบสอบถามต่อไปนี้

แบบสอบถามพื้นฐานทางสังคมและเศรษฐกิจของนักเรียน

ข้อมูลเกี่ยวกับนักเรียน

1) เพศของนักเรียน ☐ ชาย ☐ หญิง

2) อายุของนักเรียน .... ปี

ข้อมูลเกี่ยวกับครอบครัว

1) จำนวนพี่น้อง (ไม่รวมนักเรียนเอง) .... คน

2) อาชีพหลักของบิดา

☐ รับราชการ/รัฐวิสาหกิจ ☐ พนักงานเอกชน ☐ ค้าขาย/ธุรกิจ
☐ เกษตรกรรม ☐ รับจ้าง ☐ อื่น ๆ (โปรดระบุ) ....

3) อาชีพหลักของมารดา

☐ รับราชการ/รัฐวิสาหกิจ ☐ พนักงานเอกชน ☐ ค้าขาย/ธุรกิจ
☐ เกษตรกรรม ☐ รับจ้าง ☐ อื่น ๆ (โปรดระบุ) ....

4) อาชีพหลักของผู้ปกครอง (กรณีที่ผู้ดูแลหลักมิใช่บิดา/มารดาของนักเรียน)

☐ รับราชการ/รัฐวิสาหกิจ ☐ พนักงานเอกชน ☐ ค้าขาย/ธุรกิจ
☐ เกษตรกรรม ☐ รับจ้าง ☐ อื่น ๆ (โปรดระบุ) ....

5) รายได้โดยประมาณต่อเดือนของครอบครัว .... บาท

จากแบบสอบถามข้างต้น ให้นักเรียนตอบคำถามต่อไปนี้

1. จงระบุว่าประชากรและตัวอย่างของการสำรวจนี้คืออะไร (ถ้ามี)
2. จงระบุว่าตัวแปรของการสำรวจนี้คืออะไร
3. จากข้อ 2 ข้อมูลที่ได้จากตัวแปรใดบ้างเป็นข้อมูลเชิงปริมาณ และข้อมูลที่ได้จากตัวแปรใดบ้างเป็นข้อมูลเชิงคุณภาพ
4. จงระบุประเภทของข้อมูลตาม
   4.1 แหล่งที่มาของข้อมูล
   4.2 ระยะเวลาที่จัดเก็บข้อมูล
5. จงอธิบายประโยชน์ที่ได้จากการเก็บข้อมูลข้างต้น

แบบฝึกหัดท้ายบท

1. หนังสือพิมพ์ท้องถิ่นฉบับหนึ่งรายงานว่า 29% ของ 200 คน ที่สุ่มตัวอย่างมาในจังหวัดหนึ่ง ที่ได้รับการสอบถาม ระบุว่า "พวกเขาชอบไปห้างสรรพสินค้าในวันเสาร์" จงระบุว่าประชากรและตัวอย่างของการสำรวจนี้คืออะไร

2. หน่วยงานของรัฐแห่งหนึ่งต้องการประมาณสัดส่วนของเยาวชนไทยที่เชื่อว่าระบบประกันสังคมจะส่งผลดีต่อชีวิตของพวกเขาในอนาคตเมื่อพวกเขาเข้าสู่วัยชรา หน่วยงานแห่งนี้จึงได้สำรวจเยาวชนไทยจำนวน 1,500 คน และพบว่า 36% ของกลุ่มที่สำรวจมีความเชื่อดังกล่าว จงระบุว่าประชากร ตัวอย่าง และค่าสถิติของการสำรวจนี้คืออะไร

3. ในวันหนึ่งห้างสรรพสินค้าแห่งหนึ่งได้สอบถามความคิดเห็นของลูกค้าที่สุ่มตัวอย่างมาจำนวน 300 คน เกี่ยวกับความพึงพอใจต่อการให้บริการของห้างสรรพสินค้า จงระบุว่าประชากร ตัวอย่าง ตัวแปร และข้อมูลของการสำรวจนี้คืออะไร พร้อมทั้งพิจารณาว่าข้อมูลที่ห้างสรรพสินค้าจัดเก็บเป็นข้อมูลอนุกรมเวลาหรือข้อมูลตัดขวาง

4. จงพิจารณาว่าข้อมูลจากสถานการณ์ต่อไปนี้เป็นข้อมูลปฐมภูมิหรือข้อมูลทุติยภูมิ
   1) ชูใจดาวน์โหลดข้อมูลความคิดเห็นเกี่ยวกับการใช้บริการ e-Banking ของผู้ที่เข้ามาตอบแบบสอบถามออนไลน์ที่ตนสร้าง
   2) ปิติดาวน์โหลดข้อมูลราคาหุ้นจากเว็บไซต์ www.set.or.th
   3) มานีโทรศัพท์สอบถามความคิดเห็นของเพื่อน ๆ เกี่ยวกับการจัดกิจกรรมกีฬาสีของโรงเรียน

5. จงพิจารณาว่าข้อมูลต่อไปนี้เป็นข้อมูลประเภทใด (ข้อมูลอนุกรมเวลา/ข้อมูลตัดขวาง และข้อมูลเชิงปริมาณ/ข้อมูลเชิงคุณภาพ)
   1) ปริมาณกรดไขมันโอเมกา-3 ต่อเนื้อปลาทะเลไทย 100 กรัม จากปลาทั้งหมด 300 ตัวที่เก็บมา
   2) มูลค่าการส่งออกอัญมณีรายเดือนในช่วง 3 ปีที่ผ่านมา
   3) รายได้ต่อวันของร้านอาหารอิ่มอร่อยในเดือนมกราคม พ.ศ. 2562

4) ความคิดเห็นของผู้ที่อาศัยอยู่ในกรุงเทพมหานครเกี่ยวกับการจัดเก็บภาษีการรับมรดกใน พ.ศ. 2562 ว่าเห็นด้วยมากน้อยเพียงใด
5) ปริมาณการนำเข้าน้ำมันดิบรายเดือนในช่วง 3 ปีที่ผ่านมา
6) ปริมาณผลผลิตข้าวโพดจากการเก็บเกี่ยวรายปีในจังหวัดน่านตั้งแต่ พ.ศ. 2555 ถึงปัจจุบัน
7) ประเภทของสถานที่ท่องเที่ยวที่คนไทยชื่นชอบจากการสำรวจใน พ.ศ. 2562
8) ความจุผู้โดยสารของเครื่องบินโดยสารแบบต่าง ๆ
9) จำนวนผู้ป่วยโรคไส้ติ่งอักเสบของโรงพยาบาลแห่งหนึ่งใน พ.ศ. 2561
10) เชื้อชาติของชาวต่างชาติที่เดินทางเข้าประเทศไทยใน พ.ศ. 2558
11) สถานภาพสมรสของพนักงานในบริษัทแห่งหนึ่ง
12) อัตราการเกิดของประชากรไทยในแต่ละเดือนใน พ.ศ. 2562
13) อัตราการเกิดอุบัติเหตุบนทางหลวงรายเดือนในช่วง 30 ปีที่ผ่านมา
14) ร้อยละของเด็กไทยที่อ่านออกเขียนได้ใน พ.ศ. 2562
15) ยี่ห้อโทรศัพท์มือถือของนักเรียนห้องหนึ่ง
16) ระดับการศึกษาสูงสุดของอาจารย์ในมหาวิทยาลัยแห่งหนึ่ง

6 ข้อมูลนักกีฬาวอลเลย์บอลหญิงทีมชาติไทยในการแข่งขันกีฬาเอเชียนเกมส์ ค.ศ. 2014 เป็นดังนี้

เบอร์เสื้อ	ชื่อ-นามสกุล	ส่วนสูง (เซนติเมตร)	กระโดดตบสูงสุด (เซนติเมตร)	กระโดดบล็อกสูงสุด (เซนติเมตร)
2	ปิยะนุช แป้นน้อย	171	280	275
3	เอ็มอร พานุสิทธิ์	179	297	281
4	ทัดดาว นึกแจ้ง	183	305	287
5	ปลื้มจิตร์ ถินขาว	180	298	293
6	อรอุมา สิทธิรักษ์	175	303	285
9	คัทลียา ปิ่นสุวรรณ	171	291	271

เบอร์เสื้อ	ชื่อ-นามสกุล	ส่วนสูง (เซนติเมตร)	กระโดดตบสูงสุด (เซนติเมตร)	กระโดดบล็อกสูงสุด (เซนติเมตร)
10	วิลาวัณย์ อภิญญาพงศ์	174	294	282
12	ฐาปไพพรรณ ไชยศรี	168	295	276
13	นุศรา ต้อมคำ	169	289	278
15	มลิกา กันทอง	177	310	290
17	แก้วกัลยา กมุลทะลา	179	298	281
19	ปริญญา พานแก้ว	170	281	270

จากตาราง ข้อมูลในหลักใดเป็นข้อมูลเชิงปริมาณ และข้อมูลในหลักใดเป็นข้อมูลเชิงคุณภาพ

7 นักวิจัยคนหนึ่งต้องการศึกษาเกี่ยวกับภาวะเศรษฐกิจของประเทศไทยในปีที่ผ่านมา จึงสืบค้นข้อมูลจากอินเทอร์เน็ต ซึ่งได้แผนภูมิเส้นแสดง Real GDP per capita หรือผลิตภัณฑ์มวลรวมในประเทศต่อหัวที่คิดเทียบกับอัตราเงินเฟ้อของ พ.ศ. 2531 ตั้งแต่ พ.ศ. 2531 ถึง 2554 ดังนี้



จงพิจารณาว่าข้อมูลดังกล่าวเป็นข้อมูลประเภทใด (ข้อมูลปฐมภูมิ/ข้อมูลทุติยภูมิ ข้อมูลอนุกรมเวลา/ข้อมูลตัดขวาง และข้อมูลเชิงปริมาณ/ข้อมูลเชิงคุณภาพ)

เสริมสมอง : GDP

ผลิตภัณฑ์มวลรวมในประเทศ หรือ GDP (gross domestic product) คือ มูลค่าของสินค้าและบริการขั้นสุดท้ายทั้งหมดที่ผลิตได้ภายในอาณาเขตของประเทศ ในช่วงระยะเวลาหนึ่ง โดยปัจจุบันใช้เป็นตัวชี้วัดกิจกรรมทางเศรษฐกิจ และเป็นดัชนีชี้วัดที่สำคัญในการพัฒนาประเทศ อย่างไรก็ตาม การนำ GDP มาใช้เป็นตัวชี้วัดการพัฒนาประเทศมีข้อบกพร่อง เช่น GDP ไม่รวมมูลค่าสินค้าและบริการที่ไม่ผ่านตลาด แม้ว่าสินค้าและบริการดังกล่าวจะก่อให้เกิดผลผลิตเพิ่มก็ตาม GDP ไม่ได้คำนึงถึงผลกระทบทางด้านสังคมและสิ่งแวดล้อมที่เกิดจากกิจกรรมทางเศรษฐกิจ GDP ไม่ได้สะท้อนภาพสวัสดิการทางเศรษฐกิจที่แท้จริงหรือความกินดีอยู่ดีของคนส่วนใหญ่ โดยเฉพาะอย่างยิ่งถ้าสังคมนั้นหรือประเทศนั้นมีความแตกต่างในการกระจายรายได้สูง

8 จงพิจารณาว่าตัวอย่างต่อไปนี้ใช้วิธีการของสถิติศาสตร์เชิงพรรณนาหรือสถิติศาสตร์เชิงอนุมาน พร้อมทั้งให้เหตุผลประกอบ

1) สำนักโพลแห่งหนึ่งต้องการทราบความคิดเห็นของประชาชนไทยที่มีอายุ 18 – 70 ปี ทั่วประเทศต่อความสำเร็จในชีวิตของผู้หญิงในปัจจุบัน จึงได้สำรวจจากประชาชนไทยที่มีอายุ 18 – 70 ปี กระจายทุกภูมิภาคทั่วประเทศ รวมทั้งสิ้น 3,000 คน สรุปผลได้ว่าความคิดเห็นต่อความสำเร็จในชีวิตของผู้หญิงในปัจจุบันของประชาชนไทยที่มีอายุ 18 – 70 ปี ทั่วประเทศ ส่วนใหญ่ ร้อยละ 32.25 ระบุว่ามีความก้าวหน้าในหน้าที่การงาน รองลงมา ร้อยละ 24.13 ระบุว่ามีความมั่นคงทางด้านการเงิน ร้อยละ 16.77 ระบุว่าแต่งงานและมีลูก ร้อยละ 12.55 ระบุว่าเป็นเจ้าของธุรกิจหรือมีกิจการส่วนตัว ร้อยละ 8.00 ระบุว่ามีความอิสระ คล่องตัวไม่มีภาระ ร้อยละ 2.32 ระบุว่ามีโอกาสท่องเที่ยวเปิดโลกทัศน์ และร้อยละ 3.98 ระบุเหตุผลอื่น ๆ

2) กรมป้องกันและบรรเทาสาธารณภัยรายงานมูลค่าความเสียหายจากอุทกภัยใน พ.ศ. 2557 – 2560 ดังตาราง

พ.ศ.	มูลค่าความเสียหาย (บาท)
2557	323,578,804
2558	162,063,478
2559	271,167,957
2560	1,050,281,997

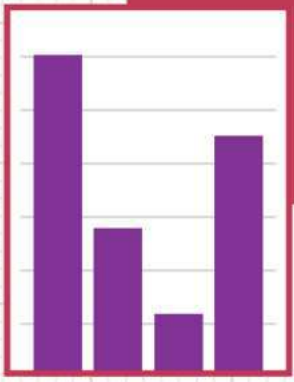
3) สำนักโพลแห่งหนึ่งต้องการทราบความคิดเห็นของประชาชนไทยที่มีอายุ 18 ปีขึ้นไป เกี่ยวกับจุดเด่นของครูไทยใน พ.ศ. 2562 จึงได้สำรวจจากประชาชนไทยที่มีอายุ 18 ปี ขึ้นไป กระจายทุกภูมิภาค ระดับการศึกษา และอาชีพ รวมทั้งสิ้น 10,000 คน สรุปผลได้ว่า ประชาชนไทยที่มีอายุ 18 ปีขึ้นไปมีความคิดเห็นว่า จุดเด่น 5 อันดับแรกของครูไทย ใน พ.ศ. 2562 ได้แก่ อันดับที่ 1 มีความทันสมัยและสามารถนำเทคโนโลยีมาใช้ในการจัดการเรียนการสอน ร้อยละ 27.13 อันดับที่ 2 ดูแลเอาใจใส่ อบรมสั่งสอนลูกศิษย์ ร้อยละ 20.22 อันดับที่ 3 เป็นแบบอย่างที่ดีให้แก่ลูกศิษย์ ร้อยละ 16.65 อันดับที่ 4 ขยัน อดทน เสียสละ ร้อยละ 9.65 และอันดับที่ 5 บุคลิกภาพ การวางตัว ร้อยละ 5.35

4) ศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมความปลอดภัยทางถนน ได้จัดทำสถิติ 10 จังหวัด ที่มีจำนวนผู้เสียชีวิตจากอุบัติเหตุบนท้องถนนมากที่สุดใน พ.ศ. 2560 แสดงดังตาราง

อันดับ	จังหวัด	จำนวนผู้เสียชีวิต (คน)
1	กรุงเทพมหานคร	845
2	ชลบุรี	522
3	นครราชสีมา	515
4	อุดรธานี	457
5	เชียงใหม่	434
6	ระยอง	425
7	สมุทรปราการ	372
8	สระบุรี	366
9	อุบลราชธานี	362
10	นครศรีธรรมราช	358

5) บริษัทผลิตรถยนต์แห่งหนึ่งกำลังจะสั่งซื้อยางรถยนต์ล็อตหนึ่งจากโรงงานผลิตยางรถยนต์ บริษัทผลิตรถยนต์แห่งนี้จึงขอสุ่มตัวอย่างยางรถยนต์มาตรวจ 1,000 เส้น ถ้าพบว่ายางรถยนต์มีตำหนิไม่เกิน 10 เส้น จะสั่งซื้อยางรถยนต์ล็อตนี้จากโรงงานผลิตยางรถยนต์

บทที่ 2 | การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพ



จุดมุ่งหมาย

1. สามารถวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่และแผนภาพ (แผนภูมิรูปภาพ แผนภูมิรูปวงกลม และแผนภูมิแท่ง) พร้อมทั้งสามารถสรุปผลที่ได้จากการนำเสนอข้อมูลด้วยตารางความถี่และแผนภาพแบบต่าง ๆ

บทที่ 2

การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพ

“

ในการดำเนินธุรกิจ ผู้ประกอบการจำเป็นต้องอาศัยข้อมูลจำนวนมากจากหลาย ๆ ด้าน เพื่อใช้ในการวางแผนการดำเนินงานและคาดการณ์ ความรู้ทางสถิติศาสตร์ในการวิเคราะห์และนำเสนอข้อมูลที่สื่อถึงคุณลักษณะต่าง ๆ ที่น่าสนใจ และการสรุปผลเพื่อสื่อความหมายที่นำไปใช้ประโยชน์ได้จึงเป็นสิ่งสำคัญ ในกรณีที่ข้อมูลที่เก็บรวบรวมมาเป็นข้อมูลเชิงคุณภาพ อาจนำเสนอข้อมูลในรูปตาราง แผนภูมิรูปภาพ แผนภูมิรูปวงกลม หรือแผนภูมิแท่ง ตัวอย่างในธุรกิจร้านขายเครื่องดื่มและของหวาน ผู้ประกอบการอาจเก็บรวบรวมข้อมูลเกี่ยวกับเพศของลูกค้าและเครื่องดื่มที่ลูกค้านิยม แล้วนำมาวิเคราะห์และนำเสนอข้อมูล เพื่อให้เข้าใจได้ง่ายและสามารถนำไปใช้ประโยชน์ต่อไปได้ โดยอาจใช้ในการสรุปประเภทเครื่องดื่มที่ลูกค้านิยมมากที่สุด หรือใช้พิจารณาว่าลูกค้าชายและหญิงนิยมเครื่องดื่มแตกต่างกันอย่างไร หรือใช้พิจารณาสัดส่วนของลูกค้าชายและหญิงที่นิยมเครื่องดื่มประเภทต่าง ๆ ผู้ประกอบการอาจนำข้อมูลที่สรุปได้เหล่านี้ไปใช้ประกอบการพิจารณาส่งข้อเสนอพิเศษให้แก่ลูกค้าแต่ละกลุ่มเพื่อจูงใจให้ลูกค้าเข้ามาใช้บริการ หรืออาจใช้ประกอบการตัดสินใจในการออกแบบผลิตภัณฑ์ใหม่ให้ตอบสนองความต้องการของลูกค้าได้มากขึ้น

”

ความรู้ก่อนหน้า

- ความหมายของสถิติศาสตร์และข้อมูล

ipst.me/10677

2.1 การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่

2.1.1 การวิเคราะห์ข้อมูลเชิงคุณภาพ

ข้อมูลเชิงคุณภาพเป็นข้อมูลที่แสดงลักษณะ ประเภท สมบัติในเชิงคุณภาพ และอื่น ๆ ที่ไม่สามารถวัดค่าเป็นตัวเลขที่นำมาบวก ลบ คูณ หรือหารกันได้ เช่น ถ้าพิจารณาตัวแปรคือเพศของผู้ใช้บริการในร้านอาหารแห่งหนึ่ง ข้อมูลที่เป็นไปได้คือหญิงหรือชาย โดยทั่วไปการเก็บข้อมูลเชิงคุณภาพที่มีจำนวนมากอาจมีข้อมูลที่ซ้ำกันอยู่ ดังนั้นจึงต้องมีการจัดระเบียบข้อมูล เพื่อให้สังเกตลักษณะของข้อมูลได้ง่ายขึ้นและสามารถหาข้อสรุปที่มีความหมายและนำไปใช้ประโยชน์ได้

ในทางสถิติศาสตร์จะวิเคราะห์ข้อมูลเชิงคุณภาพโดยพิจารณาจากความถี่และฐานนิยม ดังบทนิยามต่อไปนี้

ความถี่ (frequency) คือ จำนวนครั้งของการเกิดข้อมูลข้อมูลหนึ่งและค่าของตัวแปรค่าหนึ่ง

ฐานนิยม (mode) คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มีความถี่สูงสุดที่มากกว่า 1

ข้อมูลบางชุดอาจไม่มีฐานนิยม เช่น ในกรณีที่ข้อมูลมีความถี่เป็น 1 เท่ากันหมด นอกจากนี้ ข้อมูลบางชุดอาจมีฐานนิยมมากกว่า 1 ค่า เช่น ในกรณีที่มีข้อมูลมากกว่า 1 ข้อมูล ที่มีความถี่สูงสุดเท่ากัน

อย่างไรก็ตาม ในที่นี้จะพิจารณาเฉพาะชุดข้อมูลที่มีฐานนิยมเพียงค่าเดียว

ตัวอย่างที่ 1

โรงพยาบาลส่งเสริมสุขภาพตำบลแห่งหนึ่งได้สำรวจหมู่เลือดในระบบ ABO ของชาวบ้านที่อาศัยอยู่ใกล้โรงพยาบาลจำนวน 30 คน ได้ผลสำรวจดังนี้

A	B	O	AB	O	B	AB	O	B	A
O	O	A	B	AB	O	B	AB	O	B
O	B	O	O	A	O	A	O	B	O

จงหาความถี่ของเลือดแต่ละหมู่และฐานนิยมของข้อมูลชุดนี้

วิธีทำ จากข้อมูลที่กำหนดให้ จะได้
เลือดหมู่ A มีความถี่เป็น 5 เลือดหมู่ B มีความถี่เป็น 8
เลือดหมู่ AB มีความถี่เป็น 4 เลือดหมู่ O มีความถี่เป็น 13
จะเห็นว่า ฐานนิยมของข้อมูลชุดนี้คือ เลือดหมู่ O เนื่องจากเลือดหมู่ O มีความถี่สูงสุด ■

ตัวอย่างที่ 2

บริษัทแห่งหนึ่งได้เก็บรวบรวมข้อมูลกีฬาที่พนักงานชอบเล่นมากที่สุด โดยสำรวจจากพนักงานฝ่ายผลิตและฝ่ายขายจำนวน 60 คน ได้ข้อมูลดังนี้

พนักงานฝ่ายผลิต

แบดมินตัน	ฟุตบอล	วิ่ง	ปิงปอง	แบดมินตัน	แบดมินตัน
ฟุตบอล	แบดมินตัน	ปิงปอง	ฟุตบอล	วิ่ง	แบดมินตัน
แบดมินตัน	ปิงปอง	ปิงปอง	แบดมินตัน	ปิงปอง	ปิงปอง
วิ่ง	ปิงปอง	ฟุตบอล	แบดมินตัน	วิ่ง	วิ่ง
ฟุตบอล	วิ่ง	วิ่ง	แบดมินตัน	แบดมินตัน	ปิงปอง
ปิงปอง	แบดมินตัน	วิ่ง	ฟุตบอล	ปิงปอง	แบดมินตัน

พนักงานฝ่ายขาย

ปิงปอง	ปิงปอง	ปิงปอง	วิ่ง	ฟุตบอล	วิ่ง
แบดมินตัน	วิ่ง	ฟุตบอล	ฟุตบอล	แบดมินตัน	ปิงปอง
วิ่ง	ปิงปอง	ฟุตบอล	แบดมินตัน	วิ่ง	แบดมินตัน
ฟุตบอล	ปิงปอง	ปิงปอง	วิ่ง	ปิงปอง	ฟุตบอล

จงพิจารณาว่า

1) ถ้าบริษัทต้องการจัดการแข่งขันกีฬาเพื่อสร้างความสัมพันธ์ระหว่างพนักงานฝ่ายผลิต บริษัทควรจัดการแข่งขันกีฬาชนิดใด
2) ถ้าบริษัทต้องการจัดการแข่งขันกีฬาเพื่อสร้างความสัมพันธ์ระหว่างพนักงานฝ่ายขาย บริษัทควรจัดการแข่งขันกีฬาชนิดใด
3) ถ้าบริษัทต้องการจัดการแข่งขันกีฬาเพื่อสร้างความสัมพันธ์ระหว่างพนักงานที่สำรวจทั้งหมด บริษัทควรจัดการแข่งขันกีฬาชนิดใด

วิธีทำ 1) พิจารณาข้อมูลกีฬาที่พนักงานฝ่ายผลิตชอบเล่นมากที่สุด จะได้ว่า

วิ่ง มีความถี่เป็น 8 ฟุตบอล มีความถี่เป็น 6

ปิงปอง มีความถี่เป็น 10 แบดมินตัน มีความถี่เป็น 12

จะเห็นว่า ฐานนิยมของกีฬาที่พนักงานฝ่ายผลิตชอบเล่นมากที่สุดคือ แบดมินตัน ดังนั้น บริษัทควรจัดการแข่งขันกีฬาแบดมินตันให้กับพนักงานฝ่ายผลิต

2) พิจารณาข้อมูลกีฬาที่พนักงานฝ่ายขายชอบเล่นมากที่สุด จะได้ว่า

วิ่ง มีความถี่เป็น 6 ฟุตบอล มีความถี่เป็น 6

ปิงปอง มีความถี่เป็น 8 แบดมินตัน มีความถี่เป็น 4

จะเห็นว่า ฐานนิยมของกีฬาที่พนักงานฝ่ายขายชอบเล่นมากที่สุดคือ ปิงปอง ดังนั้น บริษัทควรจัดการแข่งขันกีฬาปิงปองให้กับพนักงานฝ่ายขาย

3) พิจารณาข้อมูลกีฬาที่พนักงานที่สำรวจทั้งหมดชอบเล่นมากที่สุด จะได้ว่า

วิ่ง มีความถี่เป็น 14 ฟุตบอล มีความถี่เป็น 12

ปิงปอง มีความถี่เป็น 18 แบดมินตัน มีความถี่เป็น 16

จะเห็นว่า ฐานนิยมของกีฬาที่พนักงานที่สำรวจทั้งหมดชอบเล่นมากที่สุดคือ ปิงปอง

ดังนั้น บริษัทควรจัดการแข่งขันกีฬาปิงปองให้กับพนักงานที่สำรวจทั้งหมด ■

ข้อมูลเชิงคุณภาพควรมีการนำเสนอข้อมูลโดยจัดเป็นหมวดหมู่ให้มีความสัมพันธ์เกี่ยวข้องกันตามสิ่งที่ต้องการทราบ โดยอาจนำเสนอในรูปตารางหรือแผนภาพ เพื่อให้สามารถอ่านและแปลความหมายของข้อมูลเหล่านั้นได้ง่ายขึ้น เรียกการนำเสนอข้อมูลโดยการแสดงความถี่ของข้อมูลในรูปตารางหรือแผนภาพว่า **การแจกแจงความถี่ (frequency distribution)**

2.1.2 การนำเสนอข้อมูลเชิงคุณภาพด้วยตารางความถี่

การนำเสนอข้อมูลเชิงคุณภาพในรูป**ตารางความถี่ (frequency table)** ที่จะกล่าวถึงในที่นี้ได้แก่ **ตารางความถี่จำแนกทางเดียว (one-way frequency table)** และ **ตารางความถี่จำแนกสองทาง (two-way frequency table)**

ตารางความถี่จำแนกทางเดียว

ตารางความถี่จำแนกทางเดียวเป็นการนำเสนอข้อมูลในรูปตารางโดยแสดงข้อมูลและความถี่ของข้อมูลของตัวแปรเพียงหนึ่งตัว มักใช้ในการนำเสนอข้อมูลเชิงคุณภาพเพื่อสรุปลักษณะที่สนใจหรือเปรียบเทียบความถี่ของแต่ละข้อมูล เช่น จากข้อมูลในตัวอย่างที่ 1 สามารถเขียนตารางความถี่จำแนกทางเดียวของข้อมูลหมู่เลือดของชาวบ้านจำนวน 30 คน ได้ดังตารางที่ 1 หรือ 2

เลือดหมู่	A	B	AB	O	**รวม**
ความถี่	5	8	4	13	**30**

ตารางที่ 1

เลือดหมู่	**ความถี่**
A	5
B	8
AB	4
O	13
รวม	**30**

ตารางที่ 2

จะเห็นว่าการนำเสนอข้อมูลด้วยตารางความถี่จำแนกทางเดียวทำให้เห็นความถี่ของข้อมูลตามลักษณะที่สนใจได้ชัดเจนกว่าการพิจารณาจากข้อมูลที่เก็บรวบรวมมาโดยตรง นอกจากนี้อาจเปรียบเทียบข้อมูลโดยใช้ความถี่สัมพัทธ์

ความถี่สัมพัทธ์ (relative frequency) คือ สัดส่วนของความถี่ของแต่ละข้อมูล เทียบกับผลรวมของความถี่ทั้งหมด

เสริมสมอง : สัดส่วนในทางสถิติศาสตร์

ในทางสถิติศาสตร์ สัดส่วน (proportion) คือ จำนวนของลักษณะหรือกลุ่มที่สนใจเมื่อเทียบกับจำนวนรวมทั้งหมด โดยกำหนดให้สัดส่วนรวมทั้งหมดเป็น 1 เช่น ในการสำรวจคนในหมู่บ้านแห่งหนึ่งจำนวน 800 คน พบว่ามีเพศหญิง 300 คน จะได้ว่าสัดส่วนของผู้อยู่อาศัยเพศหญิงในหมู่บ้านแห่งนี้ คือ $\frac{300}{800} = \frac{3}{8} = 0.375$ และสัดส่วนของผู้อยู่อาศัยเพศชายในหมู่บ้านแห่งนี้ คือ $1 - 0.375 = 0.625$

ความถี่สัมพัทธ์อาจเขียนในรูปสัดส่วน ได้เป็น

$$\text{ความถี่สัมพัทธ์ (สัดส่วน)} = \frac{\text{ความถี่}}{\text{ความถี่รวม}}$$

หรืออาจเขียนความถี่สัมพัทธ์ในรูปร้อยละ ได้เป็น

$$\text{ความถี่สัมพัทธ์ (ร้อยละ)} = \frac{\text{ความถี่}}{\text{ความถี่รวม}} \times 100$$

ในการนำเสนอข้อมูลด้วยตารางความถี่จำแนกทางเดียวอาจแสดงทั้งความถี่และความถี่สัมพัทธ์ในตารางเดียวกันได้ ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 3

จากข้อมูลในตารางที่ 2 จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ของแต่ละหมู่เลือด

วิธีทำ จากข้อมูลในตารางที่ 2 สามารถเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ได้ดังนี้

เลือดหมู่	ความถี่	ความถี่สัมพัทธ์	
		สัดส่วน	ร้อยละ
A	5	$\frac{5}{30} \approx 0.1667$	$\frac{5}{30} \times 100 \approx 16.67$
B	8	$\frac{8}{30} \approx 0.2667$	$\frac{8}{30} \times 100 \approx 26.67$
AB	4	$\frac{4}{30} \approx 0.1333$	$\frac{4}{30} \times 100 \approx 13.33$
O	13	$\frac{13}{30} \approx 0.4333$	$\frac{13}{30} \times 100 \approx 43.33$
รวม	**30**	**1**	**100**

ตัวอย่างที่ 4

ครูประจำชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง ได้สำรวจวันเกิดของนักเรียนในห้อง จำนวน 40 คน ว่ามีวันเกิดตรงกับวันใดในสัปดาห์ ได้ข้อมูลดังนี้

วันอังคาร	วันเสาร์	วันพุธ	วันอังคาร	วันจันทร์
วันอาทิตย์	วันพุธ	วันจันทร์	วันศุกร์	วันเสาร์
วันจันทร์	วันเสาร์	วันศุกร์	วันอังคาร	วันศุกร์
วันเสาร์	วันอังคาร	วันอาทิตย์	วันศุกร์	วันพุธ
วันพุธ	วันศุกร์	วันจันทร์	วันอาทิตย์	วันอาทิตย์
วันพฤหัสบดี	วันเสาร์	วันพุธ	วันอังคาร	วันพุธ
วันศุกร์	วันอังคาร	วันเสาร์	วันศุกร์	วันจันทร์
วันอังคาร	วันศุกร์	วันอังคาร	วันพุธ	วันศุกร์

จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ของข้อมูลชุดนี้ และสรุปข้อมูลที่ได้จากตาราง

วิธีทำ จากข้อมูลที่กำหนดให้ สามารถเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ ได้ดังนี้

วันเกิด	ความถี่	ความถี่สัมพัทธ์	
		สัดส่วน	ร้อยละ
วันจันทร์	5	$\frac{5}{40} = 0.125$	$0.125 \times 100 = 12.5$
วันอังคาร	8	$\frac{8}{40} = 0.2$	$0.200 \times 100 = 20$
วันพุธ	7	$\frac{7}{40} = 0.175$	$0.175 \times 100 = 17.5$
วันพฤหัสบดี	1	$\frac{1}{40} = 0.025$	$0.025 \times 100 = 2.5$
วันศุกร์	9	$\frac{9}{40} = 0.225$	$0.225 \times 100 = 22.5$
วันเสาร์	6	$\frac{6}{40} = 0.15$	$0.150 \times 100 = 15$
วันอาทิตย์	4	$\frac{4}{40} = 0.1$	$0.1 \times 100 = 10$
รวม	**40**	**1**	**100**

จากตารางสรุปได้ว่านักเรียนที่เกิดวันศุกร์มีจำนวนมากที่สุด รองลงมาคือวันอังคาร และนักเรียนที่เกิดวันพฤหัสบดีมีจำนวนน้อยที่สุด ■

ตารางความถี่จำแนกสองทาง

ตารางความถี่จำแนกสองทางเป็นการนำเสนอข้อมูลในรูปตารางซึ่งมีตัวแปรที่สนใจศึกษา 2 ตัว โดยแสดงความถี่ของข้อมูลเชิงคุณภาพของแต่ละตัวแปรที่สนใจศึกษาในรูปตาราง เช่น จากข้อมูลในตัวอย่างที่ 2 สามารถเขียนตารางความถี่จำแนกสองทางของกีฬาที่พนักงานฝ่ายผลิตและฝ่ายขายของบริษัทแห่งหนึ่งชอบเล่น ได้ดังนี้

แผนก	กีฬาที่ชอบเล่น				รวม
	วิ่ง	ฟุตบอล	ปิงปอง	แบดมินตัน	
ฝ่ายผลิต	8	6	10	12	**36**
ฝ่ายขาย	6	6	8	4	**24**
รวม	**14**	**12**	**18**	**16**	**60**

ตารางที่ 3

จากตารางสามารถสรุปได้ว่า

- พนักงานฝ่ายผลิตชอบเล่นกีฬาแบดมินตันมากที่สุด โดยคิดเป็นร้อยละ $\frac{12}{36}\times100 \approx 33.33$ ของพนักงานฝ่ายผลิตทั้งหมด
- พนักงานฝ่ายขายชอบเล่นกีฬาปิงปองมากที่สุด โดยคิดเป็นร้อยละ $\frac{8}{24}\times100 \approx 33.33$ ของพนักงานฝ่ายขายทั้งหมด
- พนักงานทั้งสองแผนกชอบเล่นกีฬาปิงปองมากที่สุด โดยคิดเป็นร้อยละ $\frac{18}{60}\times100 = 30$ ของพนักงานทั้งหมด

จากตารางข้างต้น จะเรียกความถี่ 8, 6, 10, 12, 6, 6, 8 และ 4 ว่าเป็น **ความถี่ร่วม (joint frequency)** เพราะความถี่เหล่านี้แสดงถึงจำนวนของพนักงานจากพนักงานทั้งหมด 60 คน ที่ให้ข้อมูลที่มีลักษณะร่วมกันจากทั้งสองตัวแปร เช่น ความถี่ 10 แสดงถึงจำนวนของพนักงานฝ่ายผลิตที่ชอบเล่นกีฬาปิงปอง

แบบฝึกหัด 2.1

1. ในการเลือกหัวหน้าห้องของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง ซึ่งมีนักเรียนจำนวน 30 คน โดยมีนักเรียนได้รับการเสนอชื่อเป็นหัวหน้าห้อง 3 คน คือ มานพ (M) ปรียาพร (P) และอำพล (A) ปรากฏผลดังนี้

M	M	P	M	M	P	P	M	M	P
A	M	M	A	P	A	P	A	P	M
M	A	A	A	A	M	P	M	P	P

1) จงเขียนตารางความถี่จำแนกทางเดียวพร้อมทั้งแสดงความถี่สัมพัทธ์ของข้อมูลชุดนี้
2) จงหาฐานนิยมของข้อมูลชุดนี้
3) จงสรุปเกี่ยวกับผลที่ได้จากการเลือกหัวหน้าห้องของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องนี้

2. เจ้าของร้านค้าแห่งหนึ่งได้สำรวจอาชีพของผู้มาใช้บริการที่ร้านจำนวน 60 คน โดยแบ่งตามกลุ่มอาชีพ ดังนี้
ก. ธุรกิจส่วนตัว/ค้าขาย
ข. ข้าราชการ/พนักงานรัฐวิสาหกิจ
ค. พนักงานบริษัทเอกชน
ง. นักเรียน นิสิต/นักศึกษา
จ. อาชีพอื่น ๆ
ได้ผลสำรวจดังนี้

ก	ก	ข	ก	ค	จ	ค	ง	ก	ก	ข	ข
จ	ข	ก	ก	ค	ข	จ	ก	ก	ก	ง	ค
ก	ก	ข	ข	ค	ง	ง	จ	ข	ข	ก	ก
ข	ข	ก	ก	ง	ก	ข	ค	ค	ค	ค	ง
ค	ก	ค	ข	ค	ค	ก	ข	ค	ค	ข	ก

จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ของข้อมูลชุดนี้ และสรุปข้อมูลที่ได้จากตาราง

3. จากการสำรวจของหวานที่ชื่นชอบมากที่สุดของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่ง จำนวน 315 คน โดยให้เลือกของหวานที่ชอบเพียงอย่างเดียวจากของหวาน 3 อย่าง ได้แก่ ไอศกรีม น้ำแข็งไส และเครป ได้ผลสำรวจดังนี้

เพศของนักเรียน	ของหวานที่ชื่นชอบมากที่สุด		
	ไอศกรีม	น้ำแข็งไส	เครป
ชาย	50	40	60
หญิง	65	30	70

1) นักเรียนที่ชื่นชอบน้ำแข็งไสคิดเป็นร้อยละเท่าใดของนักเรียนที่สำรวจทั้งหมด
2) นักเรียนหญิงที่ชื่นชอบไอศกรีมคิดเป็นร้อยละเท่าใดของนักเรียนหญิงที่สำรวจทั้งหมด
3) นักเรียนที่สำรวจชื่นชอบของหวานใดมากที่สุด

4. จากการสำรวจความพึงพอใจของผู้ป่วยนอกที่มีต่อการให้บริการในแผนกอายุรกรรม แผนกศัลยกรรม และแผนกทันตกรรมในโรงพยาบาลแห่งหนึ่ง โดยสำรวจจากผู้ป่วยนอกที่มาใช้บริการที่แต่ละแผนกเพียงแผนกเดียว จำนวน 220 คน ได้ผลสำรวจดังนี้

ระดับความพึงพอใจ	แผนก			
	อายุรกรรม	ศัลยกรรม	ทันตกรรม	รวม
พอใจมากที่สุด	50	20	26	**96**
พอใจมาก	22	10	18	**50**
พอใจปานกลาง	18	6	30	**54**
พอใจน้อย	6	2	4	**12**
พอใจน้อยที่สุด	4	2	2	**8**
รวม	**100**	**40**	**80**	**220**

1) ผู้ป่วยนอกที่มีระดับความพึงพอใจมากที่สุดคิดเป็นร้อยละเท่าใดของผู้ป่วยนอกที่สำรวจทั้งหมด
2) ผู้ป่วยนอกที่มีระดับความพึงพอใจน้อยที่สุดต่อการให้บริการของแผนกทันตกรรมคิดเป็นร้อยละเท่าใดของผู้ป่วยนอกที่ใช้บริการแผนกทันตกรรมที่สำรวจทั้งหมด
3) ถ้าโรงพยาบาลแห่งนี้กำหนดเกณฑ์ไว้ว่า แผนกที่ผู้ป่วยนอกมีระดับความพึงพอใจตั้งแต่พอใจมากขึ้นไป มีจำนวนมากกว่าร้อยละ 70 ของผู้ป่วยนอกที่ใช้บริการที่แผนกนั้น จะได้รับรางวัลแผนกดีเด่น จงพิจารณาว่าแผนกใดบ้างจะได้รับรางวัลนี้

2.2 การวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ

ในหัวข้อนี้ จะทบทวนเกี่ยวกับการวิเคราะห์และนำเสนอข้อมูลเชิงคุณภาพด้วยแผนภาพ ซึ่งอยู่ในรูปแผนภูมิรูปภาพ แผนภูมิรูปวงกลม และแผนภูมิแท่ง ที่ได้ศึกษามาแล้วในระดับมัธยมศึกษาตอนต้น

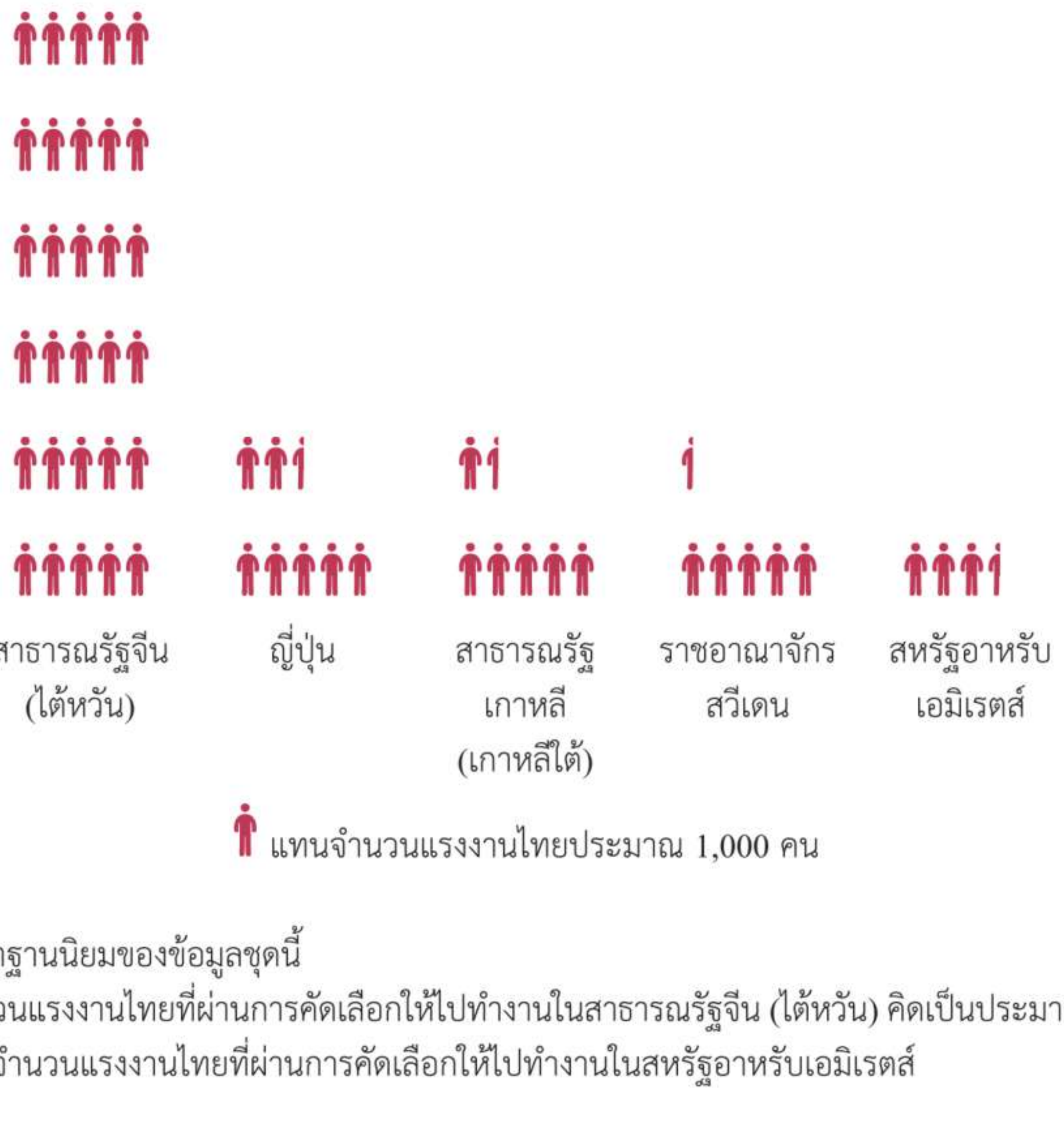
แผนภูมิรูปภาพ

แผนภูมิรูปภาพ (pictogram) เป็นการนำเสนอข้อมูลโดยใช้รูปภาพหรือสัญลักษณ์แสดงความถี่ของแต่ละข้อมูล ซึ่งจะต้องกำหนดในแผนภูมิว่ารูปภาพหรือสัญลักษณ์หนึ่งรูปนั้นแทนความถี่เท่าใด เช่น จากตารางที่ 2 สามารถเขียนแผนภูมิรูปภาพแสดงจำนวนชาวบ้านจำแนกตามหมู่เลือดในระบบ ABO ได้ดังนี้



ตัวอย่างที่ 5

จากข้อมูลของกองบริหารแรงงานไทยไปต่างประเทศ กรมการจัดหางาน กระทรวงแรงงาน พบว่า ใน พ.ศ. 2555 ประเทศที่มีแรงงานไทยผ่านการคัดเลือกให้ไปทำงานมากที่สุด 5 อันดับแรก ได้แก่ สาธารณรัฐจีน (ไต้หวัน) ญี่ปุ่น สาธารณรัฐเกาหลี (เกาหลีใต้) ราชอาณาจักรสวีเดน และสหรัฐอาหรับเอมิเรตส์ โดยจำนวนแรงงานไทยโดยประมาณที่ผ่านการคัดเลือกให้ไปทำงานในแต่ละประเทศแสดงด้วยแผนภูมิรูปภาพ ได้ดังนี้



1) จงหาฐานนิยมของข้อมูลชุดนี้
2) จำนวนแรงงานไทยที่ผ่านการคัดเลือกให้ไปทำงานในสาธารณรัฐจีน (ไต้หวัน) คิดเป็นประมาณกี่เท่าของจำนวนแรงงานไทยที่ผ่านการคัดเลือกให้ไปทำงานในสหรัฐอาหรับเอมิเรตส์

วิธีทำ 1) จากแผนภูมิรูปภาพ ฐานนิยมของข้อมูลชุดนี้คือ สาธารณรัฐจีน (ไต้หวัน) เพราะเป็นประเทศที่มีแรงงานไทยผ่านการคัดเลือกให้ไปทำงานมากที่สุด

2) เนื่องจากมีแรงงานไทยผ่านการคัดเลือกให้ไปทำงานในสาธารณรัฐจีน (ไต้หวัน) และสหรัฐอาหรับเอมิเรตส์ประมาณ 30,000 คน และ 3,500 คน ตามลำดับ
ดังนั้น จำนวนแรงงานไทยที่ผ่านการคัดเลือกให้ไปทำงานในสาธารณรัฐจีน (ไต้หวัน) คิดเป็น $\frac{30,000}{3,500} \approx 8.57$ เท่าของจำนวนแรงงานไทยที่ผ่านการคัดเลือกให้ไปทำงานในสหรัฐอาหรับเอมิเรตส์ ■

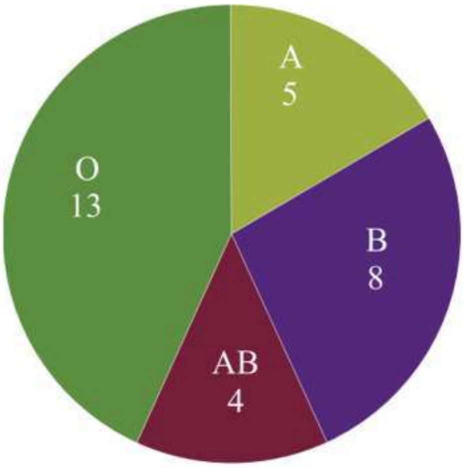
แผนภูมิรูปวงกลม

แผนภูมิรูปวงกลม (pie chart or circular chart) เป็นการนำเสนอข้อมูลโดยใช้พื้นที่ภายในของรูปวงกลมแทนความถี่ของข้อมูลทั้งหมด และแสดงสัดส่วนของความถี่ของแต่ละข้อมูลด้วยพื้นที่แต่ละส่วนภายในรูปวงกลมซึ่งแบ่งด้วยรัศมี โดยสัดส่วนของความถี่ของแต่ละข้อมูลเท่ากับสัดส่วนของขนาดของมุมที่จุดศูนย์กลางของรูปวงกลม โดยทั่วไปจะแสดงสัดส่วนของความถี่ของข้อมูลด้วยความถี่สัมพัทธ์ในรูปร้อยละ

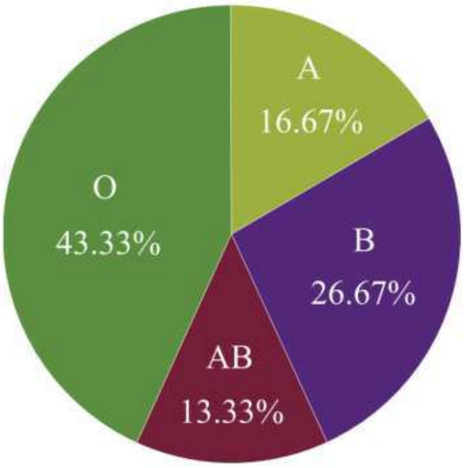
จากตารางที่ 2 สามารถคำนวณขนาดของมุมที่จุดศูนย์กลางของรูปวงกลมสำหรับความถี่ของแต่ละข้อมูลได้ดังนี้

เลือดหมู่	ความถี่	ขนาดของมุมที่จุดศูนย์กลางของรูปวงกลม
A	5	$\frac{5}{30} \times 360^\circ = 60^\circ$
B	8	$\frac{8}{30} \times 360^\circ = 96^\circ$
AB	4	$\frac{4}{30} \times 360^\circ = 48^\circ$
O	13	$\frac{13}{30} \times 360^\circ = 156^\circ$
รวม	**30**	**360°**

โดยทั่วไปไม่นิยมเขียนขนาดของมุมที่จุดศูนย์กลางของรูปวงกลมกำกับไว้ในแผนภูมิรูปวงกลม แต่จะเขียนความถี่หรือความถี่สัมพัทธ์ในรูปร้อยละของข้อมูลกำกับไว้ในแต่ละส่วน ดังนั้น จากตารางข้างต้นสามารถเขียนแผนภูมิรูปวงกลมแสดงจำนวนชาวบ้านจำแนกตามหมู่เลือดในระบบ ABO ได้ดังรูปที่ 1 และ 2



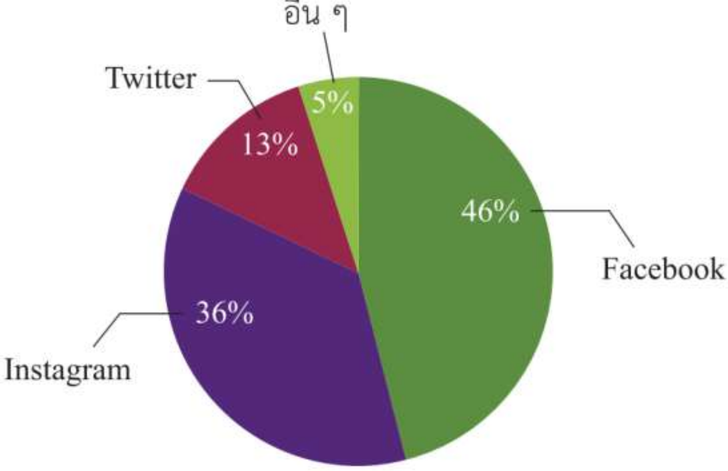
รูปที่ 1



รูปที่ 2

ตัวอย่างที่ 6

จากการสำรวจการใช้สื่อสังคมออนไลน์หลักของนักเรียน นิสิต/นักศึกษา ที่มีอายุ 15 – 25 ปี โดยผู้ตอบแบบสำรวจแต่ละคนสามารถเลือกสื่อสังคมออนไลน์หลักที่ตนเองใช้ได้เพียงสื่อเดียวเท่านั้น ได้ผลสำรวจดังนี้



1) ถ้ามีผู้ตอบแบบสำรวจทั้งหมด 1,000 คน จงหาว่ามีผู้ตอบแบบสำรวจเลือกใช้ Facebook, Instagram, Twitter และอื่น ๆ เป็นสื่อสังคมออนไลน์หลักอย่างละกี่คน
2) จงสรุปเกี่ยวกับการใช้สื่อสังคมออนไลน์หลักของนักเรียน นิสิต/นักศึกษา ที่ตอบแบบสำรวจ
3) จำนวนผู้ตอบแบบสำรวจที่ใช้ Facebook เป็นสื่อสังคมออนไลน์หลักคิดเป็นประมาณกี่เท่าของจำนวนผู้ตอบแบบสำรวจที่ใช้ Twitter เป็นสื่อสังคมออนไลน์หลัก

วิธีทำ 1) เนื่องจากมีผู้ตอบแบบสำรวจทั้งหมด 1,000 คน จะได้ว่า

มีผู้ตอบแบบสำรวจที่เลือกใช้ Facebook เป็นสื่อสังคมออนไลน์หลักจำนวน

$\frac{46}{100} \times 1,000 = 460$ คน

มีผู้ตอบแบบสำรวจที่เลือกใช้ Instagram เป็นสื่อสังคมออนไลน์หลักจำนวน

$\frac{36}{100} \times 1,000 = 360$ คน

มีผู้ตอบแบบสำรวจที่เลือกใช้ Twitter เป็นสื่อสังคมออนไลน์หลักจำนวน

$\frac{13}{100} \times 1,000 = 130$ คน

มีผู้ตอบแบบสำรวจที่เลือกใช้สื่ออื่น ๆ เป็นสื่อสังคมออนไลน์หลักจำนวน

$\frac{5}{100}\times 1,000 = 50$ คน

2) จากแผนภูมิรูปวงกลม สรุปได้ว่านักเรียน นิสิต/นักศึกษา ที่ตอบแบบสำรวจที่ใช้ Facebook เป็นสื่อสังคมออนไลน์หลักมีจำนวนมากที่สุด รองลงมาคือ Instagram, Twitter และอื่น ๆ ตามลำดับ

3) เนื่องจากมีผู้ตอบแบบสำรวจที่ใช้ Facebook และ Twitter เป็นสื่อสังคมออนไลน์หลัก คิดเป็นร้อยละ 46 และ 13 ตามลำดับ

ดังนั้น จำนวนผู้ตอบแบบสำรวจที่ใช้ Facebook เป็นสื่อสังคมออนไลน์หลักคิดเป็น

$\frac{46}{13}\approx 3.54$ เท่าของจำนวนผู้ตอบแบบสำรวจที่ใช้ Twitter เป็นสื่อสังคมออนไลน์หลัก

■

แผนภูมิแท่ง

แผนภูมิแท่ง (bar chart) เป็นการนำเสนอข้อมูลด้วยแท่งรูปสี่เหลี่ยมมุมฉากในแนวตั้งหรือแนวนอน โดยใช้ความสูงหรือความยาวของแท่งรูปสี่เหลี่ยมมุมฉากแต่ละรูปแสดงความถี่ของข้อมูลของแต่ละตัวแปรที่สนใจศึกษา

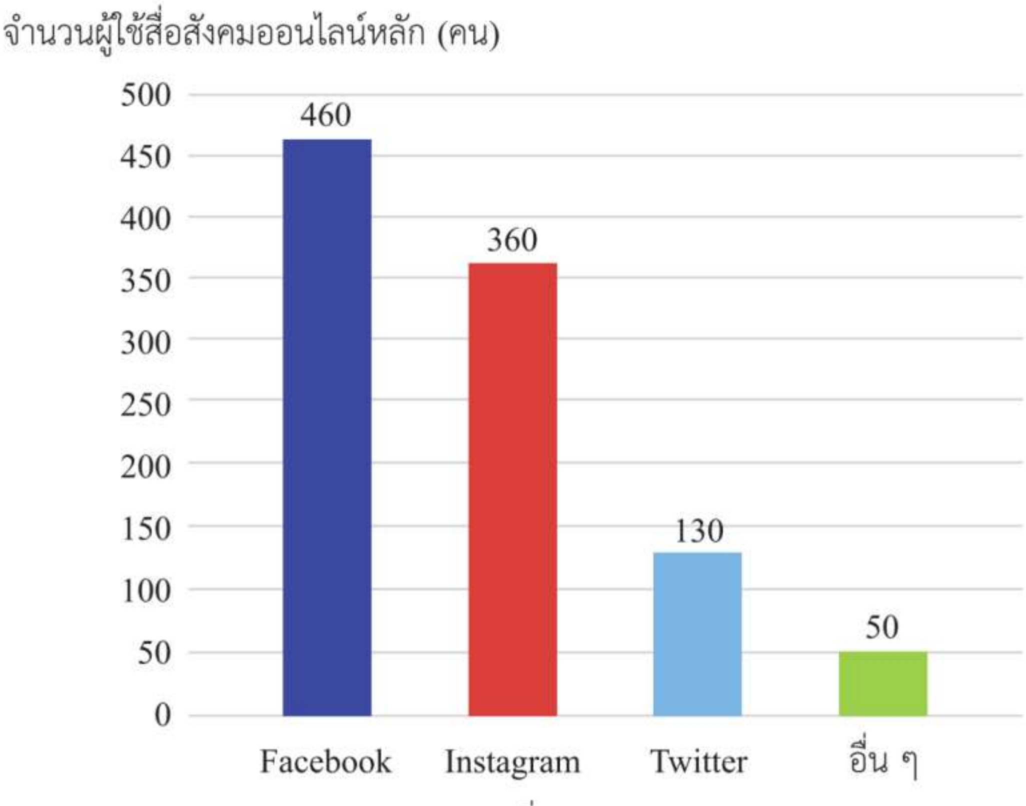
แผนภูมิแท่งมีได้หลายแบบ เช่น แผนภูมิแท่งเชิงเดี่ยว แผนภูมิแท่งพหุคูณ แผนภูมิแท่งส่วนประกอบ ในที่นี้จะทบทวนเกี่ยวกับแผนภูมิแท่งเหล่านี้โดยใช้สถานการณ์ในตัวอย่างที่ 6

1. แผนภูมิแท่งเชิงเดี่ยว

แผนภูมิแท่งเชิงเดี่ยว (simple bar chart) เป็นแผนภูมิแท่งซึ่งมีตัวแปรที่สนใจศึกษาเพียงหนึ่งตัว โดยแสดงข้อมูลและความถี่ของข้อมูล เพื่อแสดงการเปรียบเทียบความถี่ของแต่ละข้อมูล เช่น จากตัวอย่างที่ 6 ถ้ามีผู้ตอบแบบสำรวจทั้งหมด 1,000 คน จะสามารถแสดงจำนวนของผู้ใช้สื่อสังคมออนไลน์หลักแต่ละอย่าง (ความถี่) ได้ดังนี้

สื่อสังคมออนไลน์หลัก	ความถี่
Facebook	460
Instagram	360
Twitter	130
อื่น ๆ	50
รวม	**1,000**

จากตารางข้างต้น สามารถเขียนแผนภูมิแท่งเชิงเดี่ยวแสดงจำนวนของผู้ใช้สื่อสังคมออนไลน์หลักแต่ละอย่างได้ดังนี้



รูปที่ 3

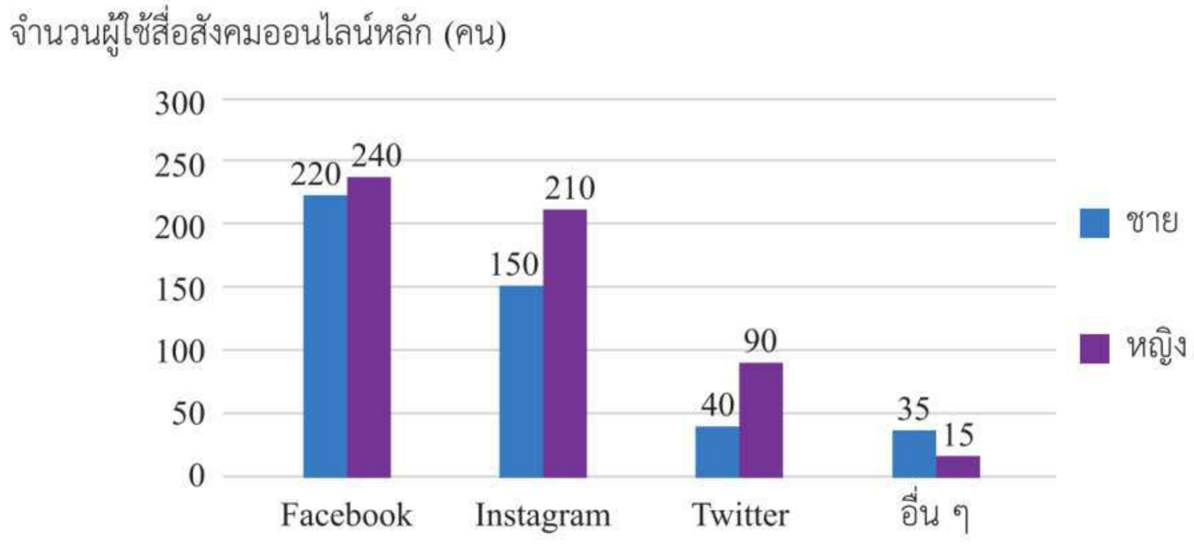
แผนภูมิแท่งเชิงเดี่ยวข้างต้นแสดงจำนวนของผู้ใช้สื่อสังคมออนไลน์หลักแต่ละอย่าง โดยเลือกอย่างใดอย่างหนึ่งจาก 4 อย่าง ได้แก่ Facebook, Instagram, Twitter และอื่น ๆ จะเห็นว่าผู้ตอบแบบสำรวจที่ใช้ Facebook เป็นสื่อสังคมออนไลน์หลักมีจำนวนมากที่สุด รองลงมาคือ Instagram, Twitter และอื่น ๆ ตามลำดับ โดยผู้ตอบแบบสำรวจที่ใช้ Twitter และอื่น ๆ เป็นสื่อสังคมออนไลน์หลักนั้น มีจำนวนน้อยกว่าผู้ตอบแบบสำรวจที่ใช้ Facebook และ Instagram เป็นสื่อสังคมออนไลน์หลักอยู่ค่อนข้างมาก

2. แผนภูมิแท่งพหุคูณ

แผนภูมิแท่งพหุคูณ (multiple bar chart) เป็นแผนภูมิแท่งซึ่งมีตัวแปรที่สนใจศึกษาตั้งแต่ 2 ตัวขึ้นไป (ในที่นี้จะศึกษาเพียง 2 ตัวเท่านั้น) โดยแสดงข้อมูลของแต่ละตัวแปรที่สนใจศึกษาบนแกนเดียวกัน และแสดงความถี่ร่วมของข้อมูลของตัวแปรที่สนใจศึกษาด้วยแท่งรูปสี่เหลี่ยมมุมฉาก เช่น จากตัวอย่างที่ 6 ถ้ากำหนดความถี่ของผู้ตอบแบบสำรวจแต่ละเพศที่เลือกใช้สื่อสังคมออนไลน์หลักแต่ละอย่าง ดังตารางต่อไปนี้

สื่อสังคมออนไลน์หลัก	เพศ		รวม
	ชาย	หญิง	
Facebook	220	240	**460**
Instagram	150	210	**360**
Twitter	40	90	**130**
อื่น ๆ	35	15	**50**
รวม	**445**	**555**	**1,000**

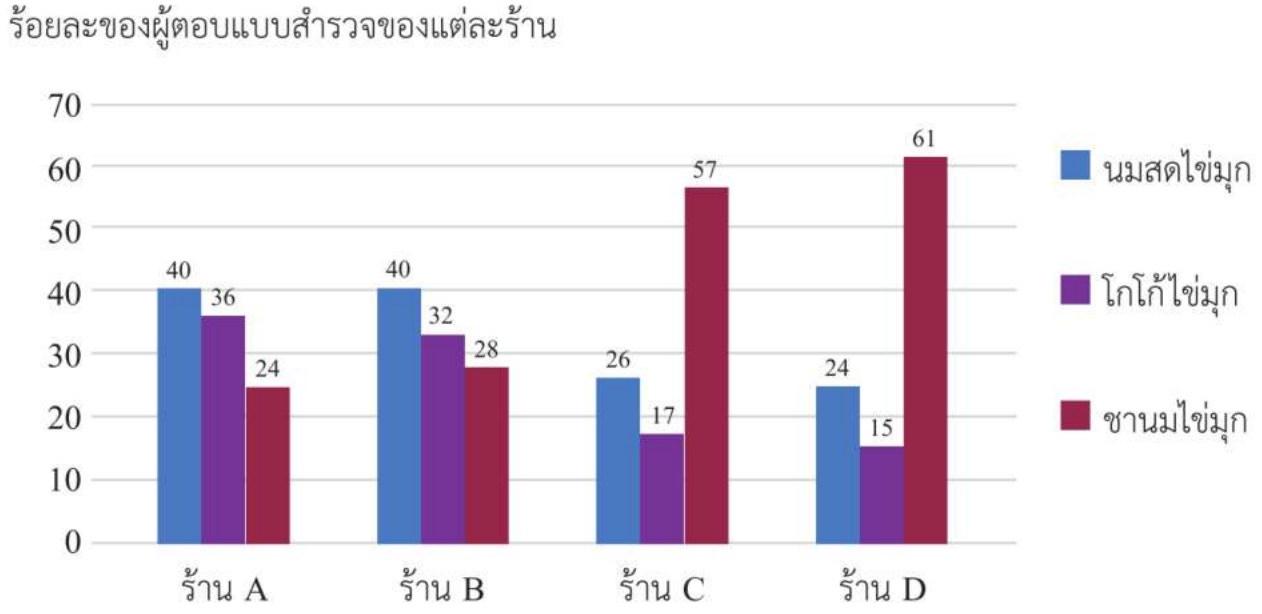
จากตารางข้างต้น สามารถเขียนแผนภูมิแท่งพหุคูณได้ดังนี้



รูปที่ 4

ตัวอย่างที่ 7

จากการสำรวจประเภทเครื่องดื่มที่ลูกค้าชื่นชอบมากที่สุด โดยมีผู้ตอบแบบสำรวจทั้งหมด 400 คน ซึ่งเป็นลูกค้าของร้าน A, B, C หรือ D ร้านใดร้านหนึ่งเพียงร้านเดียว ได้ผลสำรวจดังแผนภูมิแท่งพหุคูณต่อไปนี้



1) ถ้ามีผู้ตอบแบบสำรวจเป็นลูกค้าของแต่ละร้านจำนวนเท่ากัน จงหาว่ามีลูกค้าที่ชื่นชอบชานมไข่มุกมากที่สุดกี่คน

2) ถ้าไม่ทราบจำนวนลูกค้าของร้าน C และ D จะสรุปได้หรือไม่ว่า "จำนวนลูกค้าของร้าน C ที่ชื่นชอบโกโก้ไข่มุกมากที่สุดมีจำนวนมากกว่าจำนวนลูกค้าของร้าน D ที่ชื่นชอบโกโก้ไข่มุกมากที่สุด" เพราะเหตุใด

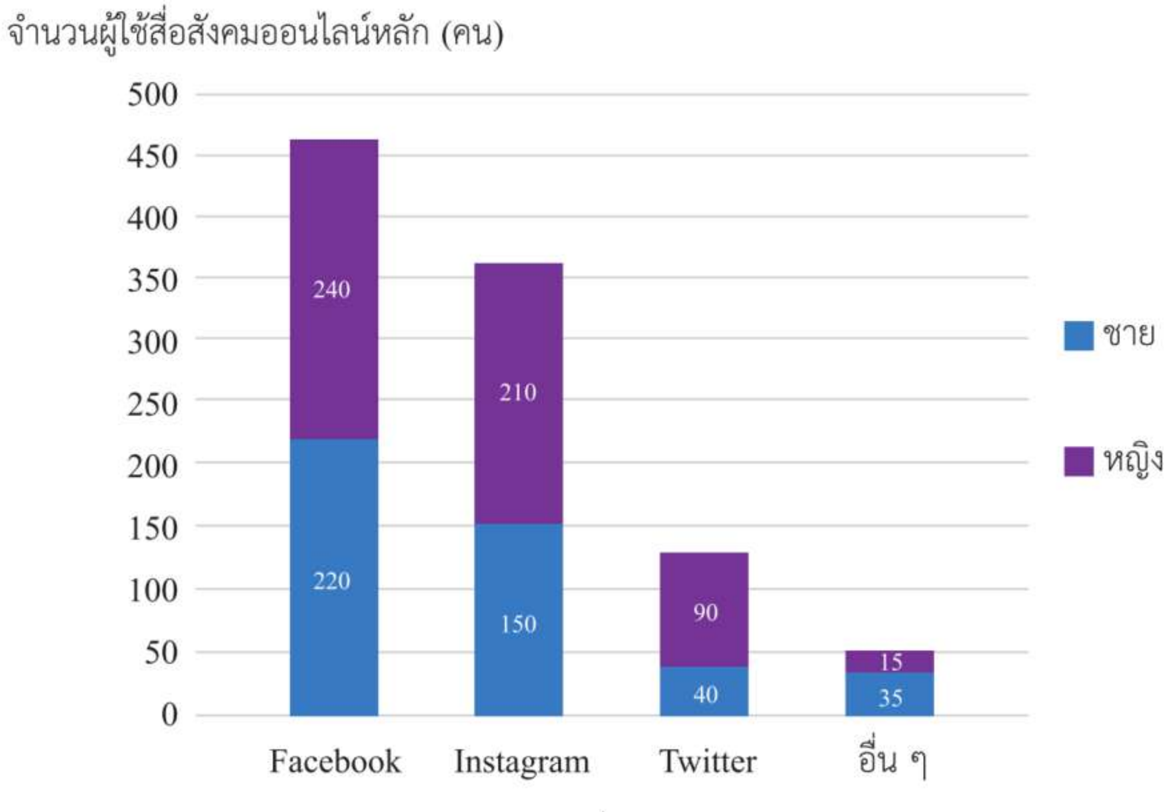
วิธีทำ 1) เนื่องจากมีผู้ตอบแบบสำรวจทั้งหมด 400 คน และมีผู้ตอบแบบสำรวจเป็นลูกค้าของแต่ละร้านจำนวนเท่ากัน จะได้ว่ามีผู้ตอบแบบสำรวจเป็นลูกค้าของแต่ละร้านจำนวน 100 คน ดังนั้น มีลูกค้าที่ชื่นชอบชานมไข่มุกมากที่สุดจำนวน $24+28+57+61=170$ คน

2) ไม่สามารถสรุปได้ เนื่องจากถ้าจำนวนลูกค้าของร้าน C และ D เท่ากัน จะได้ว่าจำนวนลูกค้าของร้าน C ที่ชื่นชอบโกโก้ไข่มุกมากที่สุดมีจำนวนมากกว่าจำนวนลูกค้าของร้าน D ที่ชื่นชอบโกโก้ไข่มุกมากที่สุด แต่ถ้าจำนวนลูกค้าของร้าน C มี 100 คน และจำนวนลูกค้าของร้าน D มี 200 คน จะได้ว่ามีลูกค้าของร้าน C ที่ชื่นชอบโกโก้ไข่มุกมากที่สุด

17 คน และมีลูกค้าของร้าน D ที่ชื่นชอบโกโก้ไข่มุกมากที่สุด 30 คน จะเห็นว่าในกรณีนี้ จำนวนลูกค้าของร้าน C ที่ชื่นชอบโกโก้ไข่มุกมากที่สุดมีจำนวนน้อยกว่าจำนวนลูกค้าของร้าน D ที่ชื่นชอบโกโก้ไข่มุกมากที่สุด ■

3. แผนภูมิแท่งส่วนประกอบ

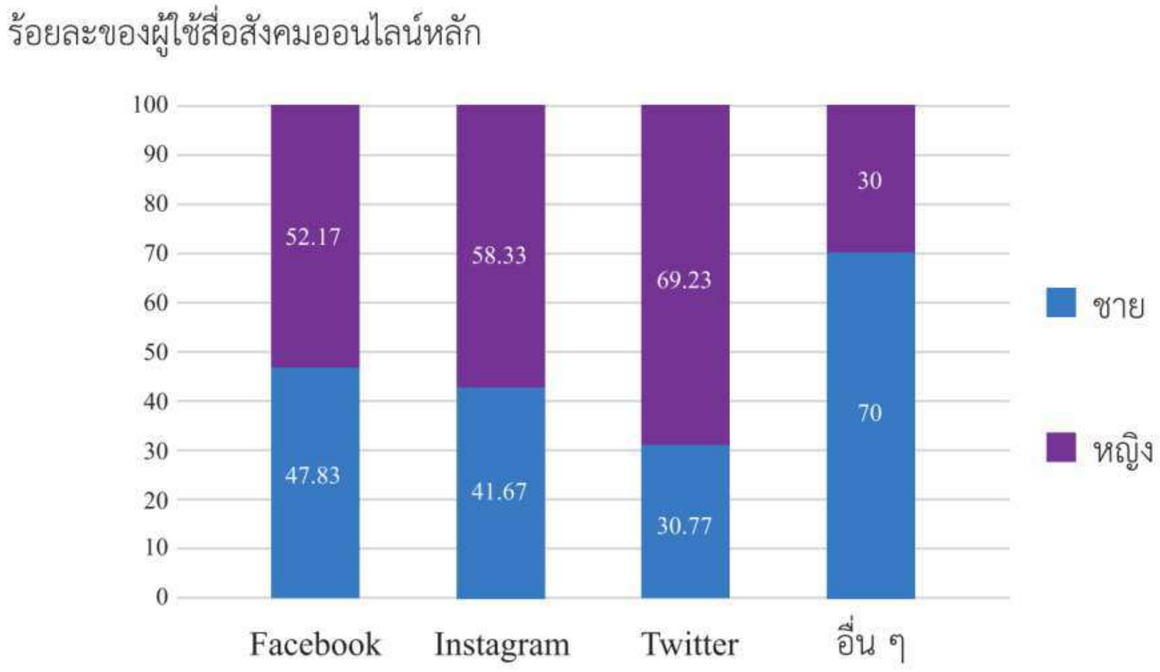
แผนภูมิแท่งส่วนประกอบ (component bar chart) เป็นแผนภูมิแท่งที่แสดงจำนวนรวมและส่วนประกอบของจำนวนรวมนั้น โดยการแบ่งเป็นส่วนย่อย ๆ เช่น จากรูปที่ 4 สามารถเขียนแผนภูมิแท่งส่วนประกอบได้โดยนำความถี่ของผู้ตอบแบบสำรวจเพศชายและหญิงที่เลือกใช้สื่อสังคมออนไลน์หลักแต่ละอย่างมาเขียนต่อเป็นแท่งเดียวกัน ดังนี้



รูปที่ 5

นอกจากจะนำเสนอข้อมูลข้างต้นด้วยแผนภูมิแท่งส่วนประกอบโดยใช้ความถี่ของแต่ละข้อมูลแล้ว ยังสามารถนำเสนอแผนภูมิแท่งส่วนประกอบโดยใช้ความถี่สัมพัทธ์ของแต่ละข้อมูล โดยจะแสดงการหาความถี่สัมพัทธ์ในรูปร้อยละของผู้ตอบแบบสำรวจแต่ละเพศที่เลือกใช้สื่อสังคมออนไลน์หลักแต่ละอย่าง และนำเสนอแผนภูมิแท่งส่วนประกอบ ได้ดังนี้

สื่อสังคมออนไลน์หลัก	เพศชาย		เพศหญิง	
	ความถี่	ความถี่สัมพัทธ์ในรูปร้อยละ	ความถี่	ความถี่สัมพัทธ์ในรูปร้อยละ
Facebook	220	$\frac{220}{460}\times100\approx 47.83$	240	$\frac{240}{460}\times100\approx 52.17$
Instagram	150	$\frac{150}{360}\times100\approx 41.67$	210	$\frac{210}{360}\times100\approx 58.33$
Twitter	40	$\frac{40}{130}\times100\approx 30.77$	90	$\frac{90}{130}\times100\approx 69.23$
อื่น ๆ	35	$\frac{35}{50}\times100=70$	15	$\frac{15}{50}\times100=30$

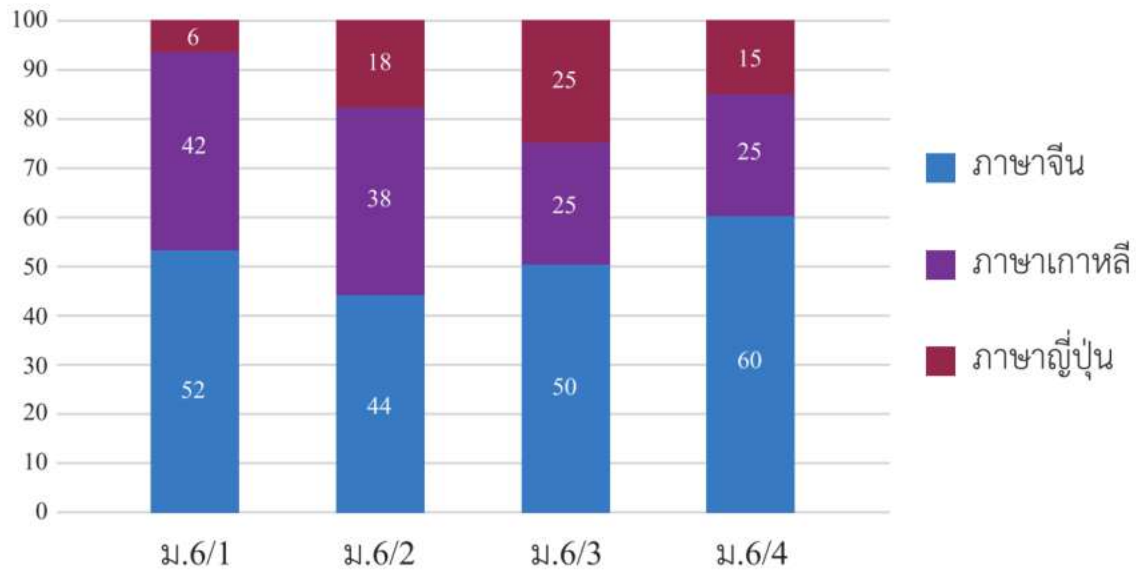


รูปที่ 6

ตัวอย่างที่ 8

จากการสำรวจความต้องการเรียนภาษาที่สามของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่ง ซึ่งมีจำนวนทั้งหมด 188 คน จาก 4 ห้อง โดยให้เลือกภาษาที่ต้องการเรียนเพียงภาษาเดียวจากภาษาจีน ภาษาเกาหลี และภาษาญี่ปุ่น ได้ผลสำรวจดังแผนภูมิแท่งส่วนประกอบต่อไปนี้

ร้อยละของนักเรียนที่ต้องการเรียนภาษาที่สาม



1) จงพิจารณาว่าโรงเรียนควรจะเปิดสอนภาษาใดให้กับนักเรียนชั้นมัธยมศึกษาปีที่ 6 จึงจะสอดคล้องกับความต้องการของนักเรียนมากที่สุด

2) ถ้าห้อง ม.6/1 และ ม.6/2 มีนักเรียนห้องละ 50 คน ห้อง ม.6/3 มีนักเรียน 48 คน และห้อง ม.6/4 มีนักเรียน 40 คน จงหาว่ามีนักเรียนที่ต้องการเรียนภาษาจีนมากหรือน้อยกว่านักเรียนที่ต้องการเรียนภาษาเกาหลีและภาษาญี่ปุ่นรวมกัน และคิดเป็นจำนวนกี่คน

วิธีทำ 1) จากแผนภูมิแท่งส่วนประกอบ จะเห็นว่าแต่ละห้องมีจำนวนนักเรียนที่ต้องการเรียนภาษาจีนมากกว่าจำนวนนักเรียนที่ต้องการเรียนภาษาอื่น ๆ
ดังนั้น โรงเรียนควรจะเปิดสอนภาษาจีนให้กับนักเรียนชั้นมัธยมศึกษาปีที่ 6 จึงจะสอดคล้องกับความต้องการของนักเรียนมากที่สุด

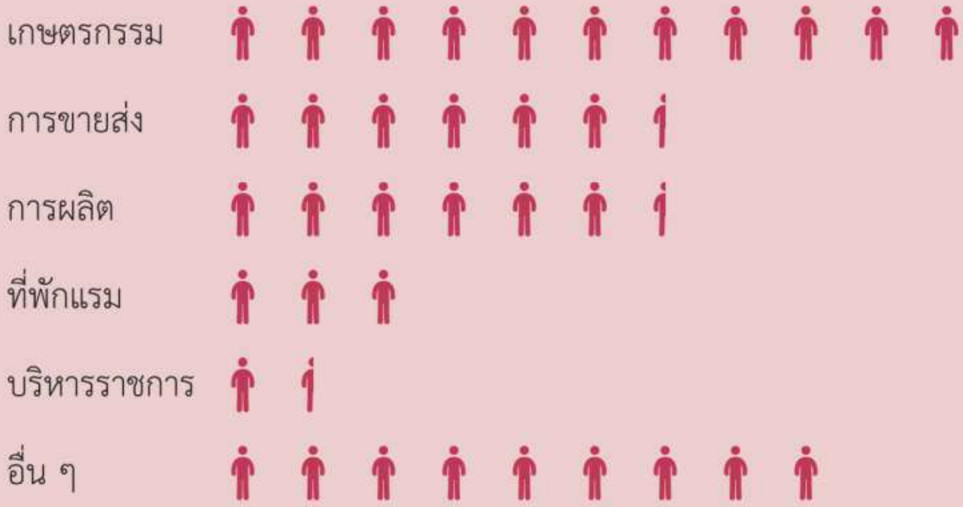
2) เนื่องจากห้อง ม.6/1 และ ม.6/2 มีนักเรียนห้องละ 50 คน ห้อง ม.6/3 มีนักเรียน 48 คน และห้อง ม.6/4 มีนักเรียน 40 คน
จะได้ว่ามีนักเรียนที่ต้องการเรียนภาษาจีนจำนวน

$$\left(\frac{52}{100}\times 50\right)+\left(\frac{44}{100}\times 50\right)+\left(\frac{50}{100}\times 48\right)+\left(\frac{60}{100}\times 40\right)=96 \text{ คน}$$

และเนื่องจากมีนักเรียนชั้นมัธยมศึกษาปีที่ 6 ทั้งหมด 188 คน จะได้ว่ามีนักเรียนที่ต้องการเรียนภาษาเกาหลีและภาษาญี่ปุ่นรวมกันจำนวน $188-96=92$ คน
ดังนั้น มีนักเรียนที่ต้องการเรียนภาษาจีนมากกว่านักเรียนที่ต้องการเรียนภาษาเกาหลีและภาษาญี่ปุ่นรวมกันจำนวน $96-92=4$ คน

แบบฝึกหัด 2.2

1. จากรายงานสรุปผลการสำรวจภาวะการทำงานของประชากร เดือนพฤษภาคม พ.ศ. 2562 จัดทำโดยสำนักงานสถิติแห่งชาติ กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม พบว่า จำนวนผู้มีงานทำจำแนกตามอุตสาหกรรมสามารถแสดงด้วยแผนภูมิรูปภาพ ได้ดังนี้



แทนจำนวนผู้มีงานทำประมาณ 1,000,000 คน

1) มีผู้ทำงานในสาขาที่พักแรมทั้งหมดประมาณกี่คน
2) จงสรุปเกี่ยวกับผู้มีงานทำจำแนกตามอุตสาหกรรม
3) จำนวนผู้ทำงานในสาขาการผลิตคิดเป็นประมาณร้อยละเท่าใดของจำนวนผู้มีงานทำทั้งหมด
4) จำนวนผู้ทำงานในสาขาเกษตรกรรมคิดเป็นประมาณกี่เท่าของจำนวนผู้ทำงานในสาขาบริหารราชการ

2. บริษัทนำเที่ยวแห่งหนึ่งได้บันทึกข้อมูลจำนวนครั้งที่จัดการนำเที่ยว ใน พ.ศ. 2562 โดยจำแนกตามภาคต่าง ๆ ของประเทศไทย ดังนี้

ภาค	จำนวนครั้งที่จัดการนำเที่ยว (ครั้ง)
เหนือ	3
ตะวันออกเฉียงเหนือ	5
กลาง	10
ตะวันออก	4
ตะวันตก	2
ใต้	1
รวม	**25**

1) จงเขียนแผนภูมิรูปภาพ แผนภูมิรูปวงกลม และแผนภูมิแท่ง แสดงจำนวนครั้งที่บริษัทแห่งนี้จัดการนำเที่ยวใน พ.ศ. 2562 จำแนกตามภาคต่าง ๆ ของประเทศไทย

2) จงหาความถี่สัมพัทธ์ในรูปสัดส่วนของจำนวนครั้งที่บริษัทแห่งนี้จัดการนำเที่ยวในแต่ละภาคของประเทศไทย

3) จำนวนครั้งที่บริษัทแห่งนี้จัดการนำเที่ยวในภาคกลางและภาคใต้รวมกันคิดเป็นร้อยละเท่าใดของจำนวนครั้งที่บริษัทแห่งนี้จัดการนำเที่ยวทั้งหมด

3. จากรายงานผลการสำรวจพฤติกรรมผู้ใช้อินเทอร์เน็ตในประเทศไทย พ.ศ. 2560 จัดทำโดยสำนักยุทธศาสตร์ สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (องค์การมหาชน) กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ในประเด็นความถี่ในการซื้อสินค้า/บริการทางออนไลน์ สามารถเขียนแผนภูมิรูปวงกลมแสดงร้อยละของผู้ใช้อินเทอร์เน็ตจำแนกตามความถี่ในการซื้อสินค้า/บริการออนไลน์ ในรอบ 3 เดือนที่ผ่านมา ได้ดังนี้

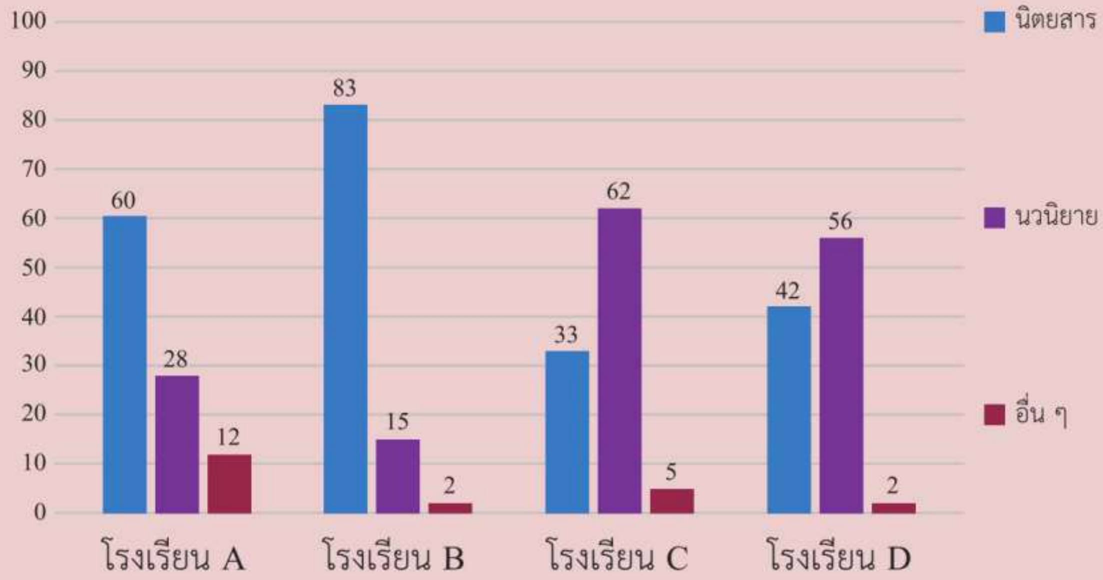


ถ้ามีผู้ใช้อินเทอร์เน็ตตอบแบบสำรวจทั้งหมด 25,000 คน จงหาว่า

1) มีผู้ใช้อินเทอร์เน็ตที่ไม่ซื้อสินค้า/บริการทางออนไลน์ ในรอบ 3 เดือนที่ผ่านมากี่คน
2) มีผู้ใช้อินเทอร์เน็ตที่ซื้อสินค้า/บริการทางออนไลน์ ในรอบ 3 เดือนที่ผ่านมากี่คน
3) จำนวนผู้ใช้อินเทอร์เน็ตที่ซื้อสินค้า/บริการทางออนไลน์เดือนละครั้งคิดเป็นกี่เท่าของจำนวนผู้ใช้อินเทอร์เน็ตที่ซื้อสินค้า/บริการทางออนไลน์มากกว่า 5 ครั้งต่อเดือน

4. จากการสำรวจประเภทของหนังสือที่นักเรียนนิยมอ่านนอกเวลาเรียนของนักเรียน 1,600 คน จาก 4 โรงเรียน โดยให้เลือกประเภทของหนังสือที่นิยมอ่านที่สุดจากนิตยสาร นวนิยาย และอื่น ๆ ได้ผลสำรวจดังแผนภูมิแท่งพหุคูณต่อไปนี้

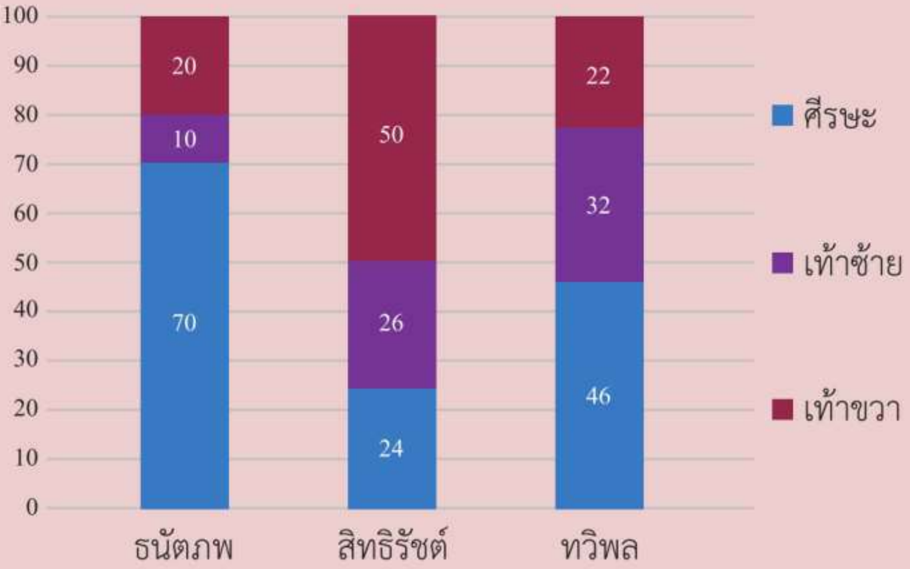
ร้อยละของนักเรียนที่นิยมอ่านหนังสือประเภทต่าง ๆ



1) ถ้ามีนักเรียนที่ตอบแบบสำรวจจากแต่ละโรงเรียนจำนวนเท่ากัน จะมีนักเรียนที่นิยมอ่านนิตยสารและนวนิยายรวมกันเท่าใด
2) ถ้ามีนักเรียนที่ตอบแบบสำรวจจากโรงเรียน A จำนวน 200 คน จากโรงเรียน B จำนวน 300 คน และจากโรงเรียน C จำนวน 600 คน จงเรียงลำดับประเภทของหนังสือที่นักเรียนนิยมอ่านจากมากที่สุดไปน้อยที่สุด

5. ผู้ฝึกสอนฟุตบอลได้เก็บข้อมูลจำนวนประตูที่ได้จากการยิงประตูด้วยศีรษะ เท้าซ้าย และเท้าขวาของนักฟุตบอล 3 คน คือ ธนัตภพ สิทธิรัชต์ และทวิพล จากการแข่งขันฟุตบอลแต่ละนัดเป็นเวลา 2 ปี ได้ผลสำรวจดังแผนภูมิแท่งส่วนประกอบต่อไปนี้

ร้อยละของจำนวนประตูที่ได้จากการยิงด้วยอวัยวะต่าง ๆ ของนักฟุตบอลแต่ละคน



1) ถ้าในเวลา 2 ปี นักฟุตบอลทั้งสามคนยิงประตูรวมกันทั้งหมด 150 ประตู โดยแต่ละคนยิงประตูได้เท่ากัน จงหาว่าจะมีจำนวนประตูที่ได้จากการยิงด้วยเท้าทั้งหมดเท่าใด
2) ถ้าในเวลา 2 ปี นักฟุตบอลทั้งสามคนยิงประตูรวมกันทั้งหมด 200 ประตู โดยสิทธิรัชต์และธนัตภพยิงประตูได้เท่ากัน ส่วนทวิพลยิงประตูได้สองเท่าของสิทธิรัชต์ จงพิจารณาว่าจำนวนประตูที่สิทธิรัชต์ยิงด้วยเท้าขวามากหรือน้อยกว่าจำนวนประตูที่ทวิพลยิงด้วยเท้าซ้าย
3) ถ้านักฟุตบอลทั้งสามคนยิงประตูได้เท่ากัน จงพิจารณาว่าจะสามารถสรุปข้อความต่อไปนี้ได้หรือไม่ เพราะเหตุใด
   3.1) ในจำนวนประตูที่ได้จากการยิงด้วยศีรษะ สิทธิรัชต์ยิงประตูได้น้อยที่สุด เมื่อเทียบกับนักฟุตบอลคนอื่น ๆ
   3.2) ประตูที่ได้จากการยิงด้วยเท้าซ้ายมีจำนวนน้อยที่สุด เมื่อเทียบกับจำนวนประตูที่ได้จากการยิงด้วยอวัยวะอื่น ๆ
4) ถ้าไม่ทราบจำนวนประตูที่นักฟุตบอลแต่ละคนยิงได้ จงพิจารณาว่าจะสามารถสรุปข้อความในข้อ 3.1) ได้หรือไม่ เพราะเหตุใด

กิจกรรม : เสื้อรุ่นที่ระลึก

นักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 50 คน เก็บข้อมูลขนาดเสื้อยืดเพื่อสั่งทำเสื้อรุ่นที่ระลึกเนื่องในโอกาสจบการศึกษา ได้ข้อมูลดังนี้

S	M	L	XL	L	XL	M	S	L	L
XS	M	L	M	S	M	L	M	XL	M
S	S	S	S	XS	M	M	L	L	S
S	XS	S	XS	S	M	M	L	XS	S
XL	L	M	M	M	S	XL	L	S	M

ขั้นตอนการปฏิบัติ

1. นำเสนอข้อมูลดังกล่าวด้วยรูปแบบการนำเสนอที่คิดว่าเหมาะสม พร้อมทั้งให้เหตุผลประกอบ
2. สรุปเกี่ยวกับขนาดเสื้อยืดของนักเรียนห้องนี้
3. ถ้านักเรียนห้องนี้เป็นชาย 24 คน และหญิง 26 คน โดยตารางความถี่จำแนกสองทางของขนาดเสื้อยืดของนักเรียนชายและนักเรียนหญิงเป็นดังนี้

เพศ	ขนาดเสื้อยืด				
	XS	**S**	**M**	**L**	**XL**
ชาย	1	5	8	7	3
หญิง	4	9	7	4	2

3.1) นำเสนอข้อมูลข้างต้นด้วยรูปแบบการนำเสนอที่คิดว่าเหมาะสม พร้อมทั้งให้เหตุผลประกอบ
3.2) สรุปเกี่ยวกับขนาดเสื้อยืดของนักเรียนชายและนักเรียนหญิงห้องนี้

แบบฝึกหัดท้ายบท

1 จากการสำรวจภูมิลำเนาของพนักงานในบริษัทแห่งหนึ่งจำนวน 60 คน พบว่าพนักงานเหล่านี้มาจากประเทศไทย (T) สาธารณรัฐประชาชนจีน (C) สาธารณรัฐสิงคโปร์ (S) หรือสาธารณรัฐสังคมนิยมเวียดนาม (V) ได้ผลสำรวจดังนี้

T	V	S	T	C	V	T	T	C	S
S	T	C	C	S	T	T	V	T	T
V	V	S	V	T	S	T	T	S	T
S	T	C	T	C	T	C	V	C	T
C	C	T	V	T	V	S	T	S	S
T	V	S	T	S	T	V	V	C	T

1) จงเขียนตารางความถี่จำแนกทางเดียวพร้อมทั้งแสดงความถี่สัมพัทธ์ของข้อมูลชุดนี้
2) จงหาฐานนิยมของข้อมูลชุดนี้
3) พนักงานที่มาจากประเทศในเอเชียตะวันออกเฉียงใต้คิดเป็นร้อยละเท่าใดของพนักงานทั้งหมด
4) จงสรุปเกี่ยวกับภูมิลำเนาของพนักงานในบริษัทแห่งนี้

2 จากการสำรวจเกี่ยวกับระดับราคาที่ลูกค้าซื้อคอมพิวเตอร์ในร้านค้าแห่งหนึ่ง โดยมีคำถาม 2 ข้อ ดังนี้

1. เพศ ☐ ชาย ☐ หญิง
2. ระดับราคาที่ซื้อคอมพิวเตอร์
   ☐ ระดับราคาที่ 1: ต่ำกว่า 35,001 บาท
   ☐ ระดับราคาที่ 2: 35,001 – 45,000 บาท
   ☐ ระดับราคาที่ 3: 45,001 – 55,000 บาท
   ☐ ระดับราคาที่ 4: 55,001 – 65,000 บาท
   ☐ ระดับราคาที่ 5: 65,001 – 75,000 บาท
   ☐ ระดับราคาที่ 6: สูงกว่า 75,000 บาท

ได้ผลสำรวจดังตารางความถี่ต่อไปนี้

ระดับราคา	ชาย		หญิง		รวม	
	จำนวน	ร้อยละ	จำนวน	ร้อยละ	จำนวน	ร้อยละ
ระดับราคาที่ 1	58	45.7	0	0.0	58	29.0
ระดับราคาที่ 2	12	9.4	40	54.8	52	26.0
ระดับราคาที่ 3	24	18.9	0	0.0	24	12.0
ระดับราคาที่ 4	0	0.0	17	23.3	17	8.5
ระดับราคาที่ 5	17	13.4	0	0.0	17	8.5
ระดับราคาที่ 6	16	12.6	16	21.9	32	16.0
รวม	**127**	**100.0**	**73**	**100.0**	**200**	**100.0**

จงเขียนแผนภูมิแท่งส่วนประกอบเพื่อนำเสนอข้อมูลชุดนี้

3 จากการสำรวจนักเรียนกลุ่มหนึ่งจำนวน 12 คน โดยสอบถามเพศและบุคคลที่นักเรียนพักอาศัยด้วย ได้ข้อมูลดังนี้

คนที่	เพศ	บุคคลที่พักอาศัยด้วย
1	ชาย	บิดา/มารดา
2	ชาย	ญาติ
3	หญิง	ญาติ
4	หญิง	บิดา/มารดา
5	หญิง	ญาติ
6	หญิง	ญาติ

คนที่	เพศ	บุคคลที่พักอาศัยด้วย
7	ชาย	บิดา/มารดา
8	ชาย	ญาติ
9	ชาย	ญาติ
10	หญิง	บิดา/มารดา
11	ชาย	ญาติ
12	ชาย	บิดา/มารดา

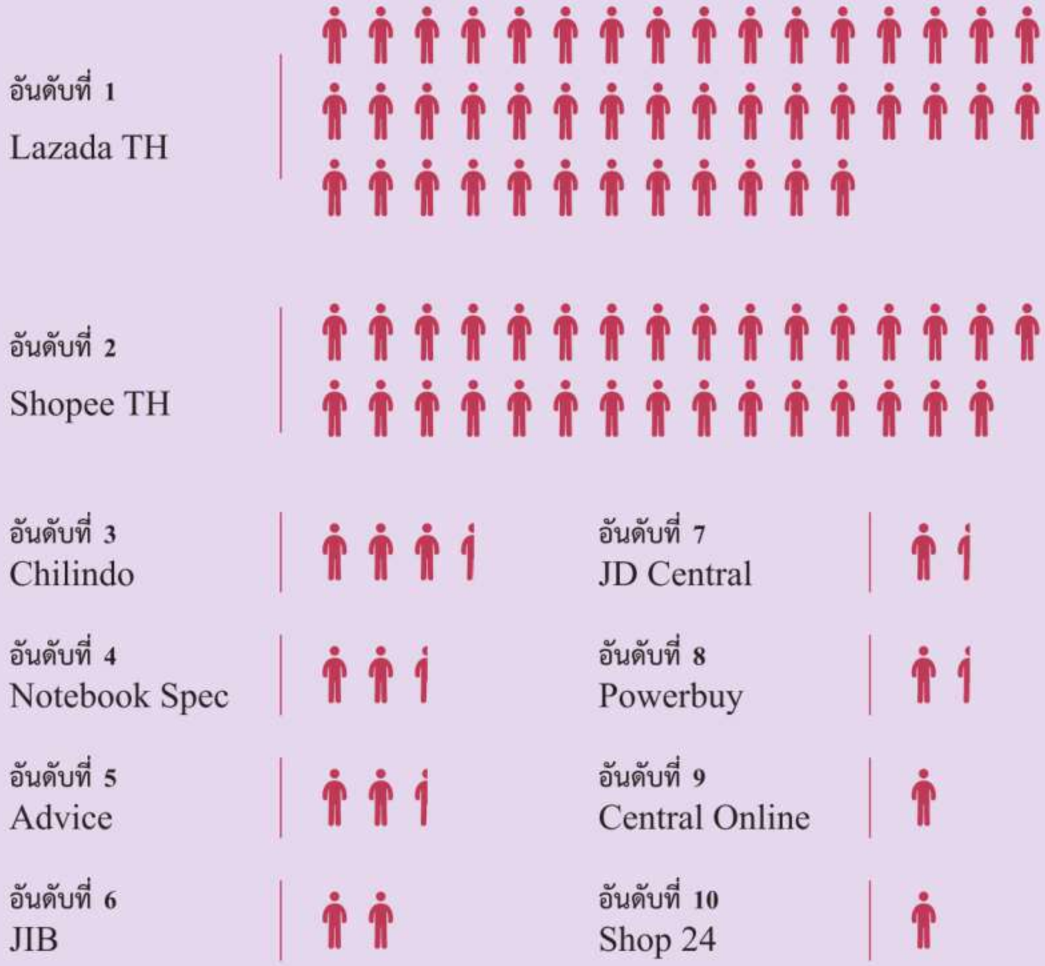
1) จงเขียนตารางความถี่จำแนกทางเดียวโดยจำแนกตามเพศ
2) จงเขียนตารางความถี่จำแนกทางเดียวโดยจำแนกตามบุคคลที่พักอาศัยด้วย
3) จงเขียนตารางความถี่จำแนกสองทางเพื่อนำเสนอข้อมูลชุดนี้
4) นักเรียนที่สำรวจพักอาศัยกับบิดา/มารดาหรือญาติมากกว่ากัน

4 จากการสำรวจผลไม้ที่ชื่นชอบของนักเรียนและครูของโรงเรียนแห่งหนึ่งจำนวน 245 คน โดยให้เลือกผลไม้ที่ชื่นชอบเพียง 1 ชนิด จากแตงโม มะม่วง ฝรั่ง และสับปะรด ได้ผลสำรวจดังนี้

กลุ่มที่ตอบแบบสำรวจ	ผลไม้ที่ชื่นชอบ				รวม
	แตงโม	มะม่วง	ฝรั่ง	สับปะรด	
นักเรียนระดับประถมศึกษาตอนต้น	20	10	10	5	**45**
นักเรียนระดับประถมศึกษาตอนปลาย	12	14	15	10	**51**
นักเรียนระดับมัธยมศึกษาตอนต้น	11	13	17	15	**56**
นักเรียนระดับมัธยมศึกษาตอนปลาย	10	12	12	14	**48**
ครู	7	16	8	14	**45**
รวม	**60**	**65**	**62**	**58**	**245**

1) นักเรียนที่ชื่นชอบมะม่วงคิดเป็นร้อยละเท่าใดของนักเรียนที่สำรวจทั้งหมด
2) นักเรียนและครูที่ชื่นชอบแตงโมหรือมะม่วงคิดเป็นร้อยละเท่าใดของนักเรียนและครูที่สำรวจทั้งหมด
3) จงเรียงลำดับผลไม้ที่นักเรียนชื่นชอบจากมากที่สุดไปน้อยที่สุด
4) ถ้าโรงเรียนแห่งนี้จะให้ขายผลไม้ที่มีนักเรียนและครูชื่นชอบมากกว่าร้อยละ 25 ของนักเรียนและครูที่สำรวจทั้งหมด จงหาว่าโรงเรียนแห่งนี้จะมีผลไม้ชนิดใดขายบ้าง

5 จากการสำรวจจำนวนผู้เข้าชมเว็บไซต์ร้านค้า e-Commerce โดย Map of e-Commerce ในเดือนเมษายน พ.ศ. 2562 ได้แผนภูมิรูปภาพแสดง 10 อันดับของเว็บไซต์ร้านค้า e-Commerce ที่มีผู้เข้าชมมากที่สุดในประเทศไทย ดังนี้



แทนจำนวนผู้เข้าชมเว็บไซต์ร้านค้า e-Commerce ประมาณ 1,000,000 คน

1) เว็บไซต์ร้านค้า e-Commerce สองอันดับแรกมีจำนวนผู้เข้าชมต่างกันประมาณเท่าใด
2) จำนวนผู้เข้าชมเว็บไซต์ Shopee TH คิดเป็นประมาณกี่เท่าของจำนวนผู้เข้าชมเว็บไซต์ Shop 24

เสริมสมอง : e-Commerce

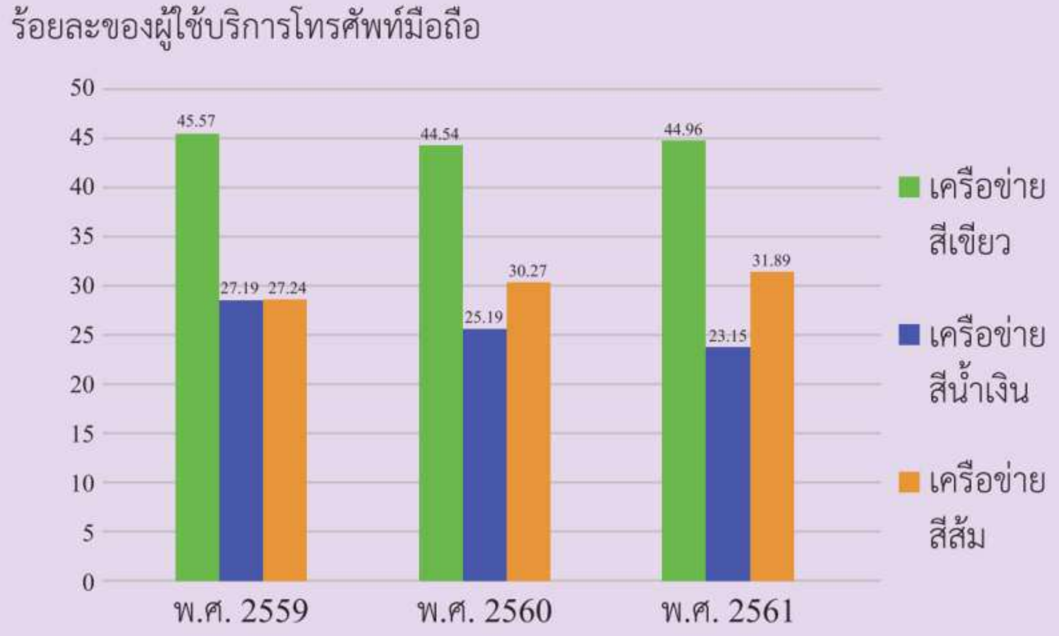
พาณิชย์อิเล็กทรอนิกส์ หรือ e-Commerce (electronic commerce) คือ การทำธุรกรรมซื้อขาย หรือแลกเปลี่ยนสินค้าและบริการบนอินเทอร์เน็ต โดยใช้เว็บไซต์หรือแอปพลิเคชันเป็นสื่อในการนำเสนอสินค้าและบริการต่าง ๆ รวมถึงการติดต่อกันระหว่างผู้ซื้อและผู้ขาย ทำให้ผู้เข้าใช้บริการจากทุกที่ทุกประเทศ หรือทุกมุมโลกสามารถเข้าถึงร้านค้าได้ง่ายและตลอด 24 ชั่วโมง

6 จากการสำรวจขนาดเสื้อยืดของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 40 คน ได้ข้อมูลดังนี้

ขนาดเสื้อยืด	จำนวนนักเรียน
XS	2
S	8
M	12
L	10
XL	6
XXL	2
รวม	**40**

1) จงเขียนแผนภูมิรูปภาพ แผนภูมิรูปวงกลม และแผนภูมิแท่งเพื่อนำเสนอข้อมูลชุดนี้
2) จงหาความถี่สัมพัทธ์ในรูปร้อยละของจำนวนนักเรียนที่ใส่เสื้อยืดขนาดต่าง ๆ
3) นักเรียนที่ใส่เสื้อยืดตั้งแต่ขนาด L ขึ้นไป คิดเป็นร้อยละเท่าใดของนักเรียนที่สำรวจทั้งหมด

7 จากการสำรวจจำนวนผู้ใช้บริการโทรศัพท์มือถือใน พ.ศ. 2559 – 2561 โดยจำแนกตามเครือข่ายที่ให้บริการ ได้ผลสำรวจดังแผนภูมิแท่งพหุคูณต่อไปนี้



1) เครือข่ายใดมีจำนวนผู้ใช้บริการมากเป็นอันดับที่สองตลอดทั้งสามปีที่ทำการสำรวจ

2) ถ้าใน พ.ศ. 2560 มีจำนวนผู้ใช้บริการโทรศัพท์มือถือเครือข่ายสีน้ำเงิน 24,480,000 คน จงหาจำนวนผู้ใช้บริการโทรศัพท์มือถือทั้งหมดใน พ.ศ. 2560

3) ถ้าใน พ.ศ. 2559 มีจำนวนผู้ใช้บริการโทรศัพท์มือถือทั้งหมด 90,000,000 คน และในปีต่อ ๆ ไป มีจำนวนผู้ใช้บริการเพิ่มขึ้นปีละ 1% ของปีก่อนหน้า จงหาว่าใน พ.ศ. 2561 มีจำนวนผู้ใช้บริการโทรศัพท์มือถือเครือข่ายสีส้มกับสีเขียวต่างกันประมาณกี่คน

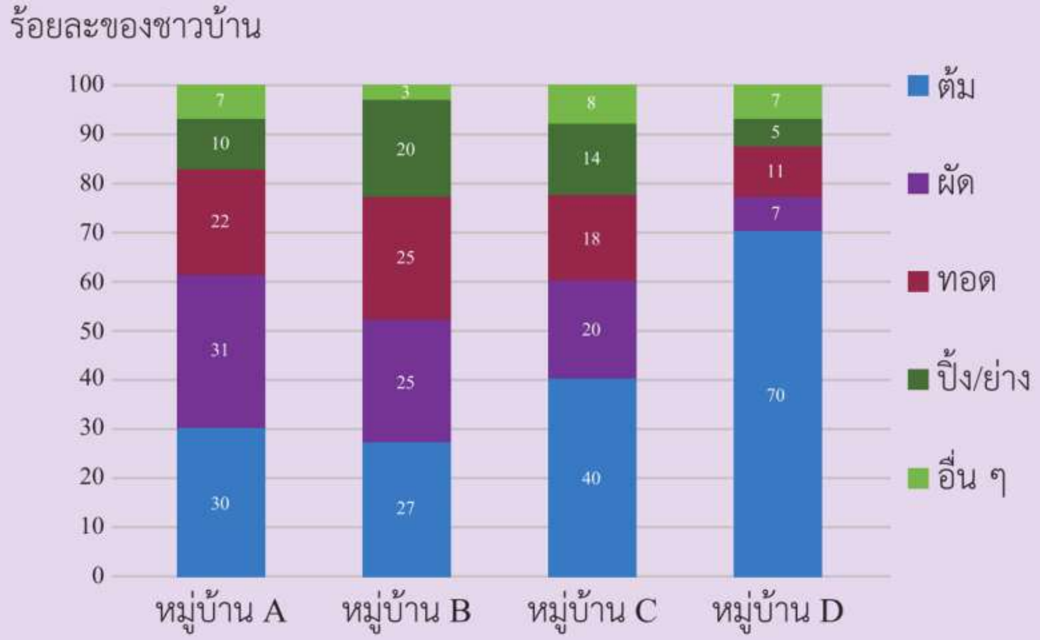
4) จงพิจารณาว่าข้อความต่อไปนี้เป็นจริง เป็นเท็จ หรือไม่สามารถสรุปได้ว่าเป็นจริงหรือเท็จ พร้อมทั้งให้เหตุผลประกอบ

4.1) เครือข่ายสีเขียวมีจำนวนผู้ใช้บริการมากที่สุดตลอดทั้งสามปีที่ทำการสำรวจ

4.2) จำนวนผู้ใช้บริการโทรศัพท์มือถือเครือข่ายสีส้มเพิ่มขึ้นทุกปี

4.3) ถ้าใน พ.ศ. 2559 มีจำนวนผู้ใช้บริการโทรศัพท์มือถือทั้งหมด 90,000,000 คน และในปีต่อ ๆ ไป มีจำนวนผู้ใช้บริการเพิ่มขึ้นปีละ 1% ของปีก่อนหน้า แล้วจำนวนผู้ใช้บริการโทรศัพท์มือถือเครือข่ายสีเขียวใน พ.ศ. 2561 น้อยกว่าจำนวนผู้ใช้บริการเครือข่ายสีเขียวใน พ.ศ. 2559

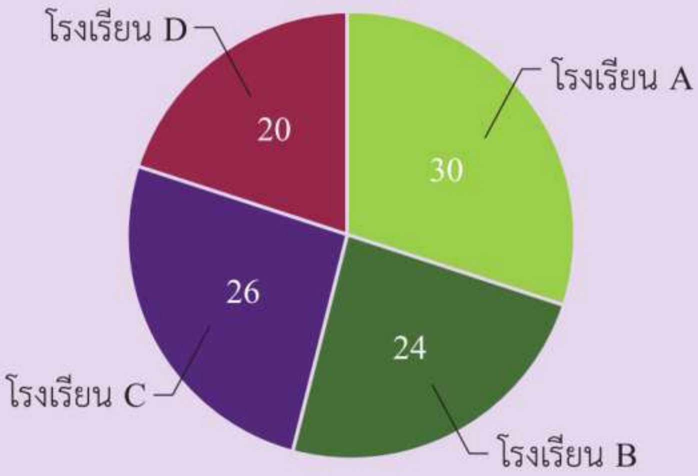
8 จากการสำรวจชาวบ้านจากหมู่บ้าน 4 แห่ง จำนวน 800 คน เกี่ยวกับวิธีปรุงอาหารที่รับประทานเป็นประจำ โดยให้เลือกเพียง 1 วิธี จากการต้ม ผัด ทอด ปิ้ง/ย่าง และวิธีอื่น ๆ ได้ผลสำรวจดังแผนภูมิแท่งส่วนประกอบต่อไปนี้



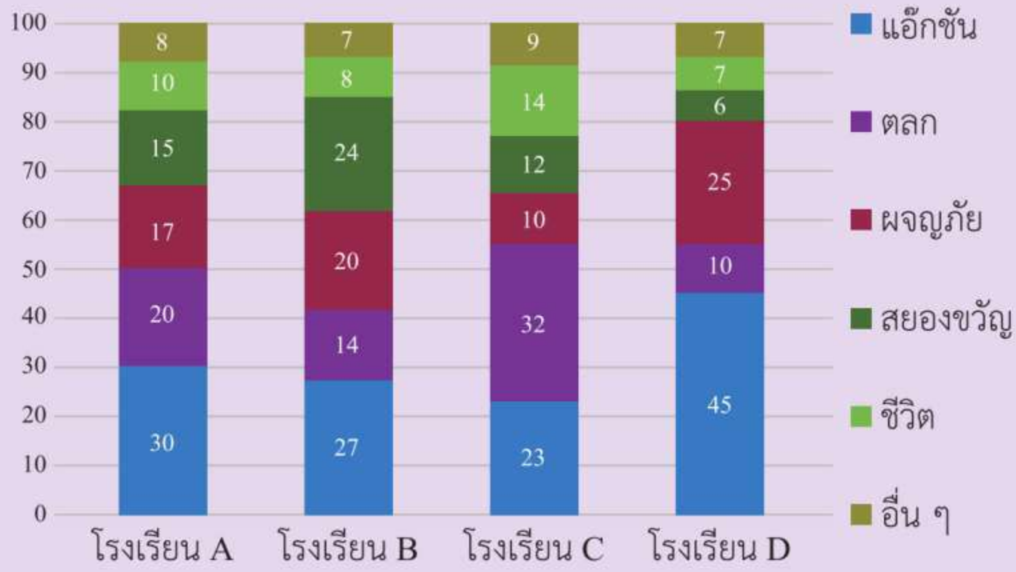
1) ถ้าแต่ละหมู่บ้านมีชาวบ้านที่ตอบแบบสำรวจจำนวนเท่ากัน จงหาจำนวนชาวบ้านที่รับประทานอาหารที่ปรุงด้วยการทอดเป็นประจำ
2) ถ้ามีชาวบ้านที่ตอบแบบสำรวจจากหมู่บ้าน A จำนวน 100 คน จากหมู่บ้าน B จำนวน 150 คน และจากหมู่บ้าน D จำนวน 50 คน จงพิจารณาว่าจำนวนชาวบ้านในหมู่บ้าน A ที่รับประทานอาหารที่ปรุงด้วยการผัดเป็นประจำมากหรือน้อยกว่าจำนวนชาวบ้านในหมู่บ้าน C ที่รับประทานอาหารที่ปรุงด้วยการผัดเป็นประจำ
3) จงพิจารณาว่าข้อความต่อไปนี้เป็นจริง เป็นเท็จ หรือไม่สามารถสรุปได้ว่าเป็นจริงหรือเท็จ พร้อมทั้งให้เหตุผลประกอบ
   3.1) ชาวบ้านที่รับประทานอาหารที่ปรุงด้วยการทอดเป็นประจำมีจำนวนน้อยกว่าชาวบ้านที่รับประทานอาหารที่ปรุงด้วยการต้มเป็นประจำ
   3.2) ชาวบ้านในหมู่บ้าน B ที่รับประทานอาหารที่ปรุงด้วยการผัดเป็นประจำมีจำนวนมากกว่าชาวบ้านในหมู่บ้าน C ที่รับประทานอาหารที่ปรุงด้วยการปิ้ง/ย่างเป็นประจำ
   3.3) ชาวบ้านที่รับประทานอาหารที่ปรุงด้วยการทอดเป็นประจำมีจำนวนน้อยกว่าชาวบ้านที่รับประทานอาหารที่ปรุงด้วยการผัดเป็นประจำ

9 จากการสำรวจประเภทของภาพยนตร์ที่นักเรียนชื่นชอบ โดยสำรวจจากนักเรียนจำนวน 10,000 คน จาก 4 โรงเรียน และให้เลือกประเภทของภาพยนตร์เพียงประเภทเดียว จากภาพยนตร์แอ๊กชัน ตลก ผจญภัย สยองขวัญ ชีวิต และภาพยนตร์ประเภทอื่น ๆ ได้ผลสำรวจดังแผนภูมิรูปวงกลมและแผนภูมิแท่งส่วนประกอบต่อไปนี้

ร้อยละของนักเรียนที่สำรวจจากแต่ละโรงเรียน



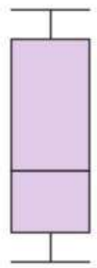
ร้อยละของนักเรียนที่ชื่นชอบภาพยนตร์แต่ละประเภท



จากแผนภูมิรูปวงกลมและแผนภูมิแท่งส่วนประกอบข้างต้น จงตอบคำถามต่อไปนี้

1) มีนักเรียนจากโรงเรียน A ที่ชื่นชอบภาพยนตร์แอ็กชันกี่คน
2) มีนักเรียนที่ชื่นชอบภาพยนตร์ผจญภัยกี่คน
3) จำนวนนักเรียนจากโรงเรียน A ที่ชื่นชอบภาพยนตร์ตลกมากหรือน้อยกว่าจำนวนนักเรียนจากโรงเรียน D ที่ชื่นชอบภาพยนตร์ผจญภัย
4) มีนักเรียนจากโรงเรียน B และ C ที่ชื่นชอบภาพยนตร์ผจญภัยและภาพยนตร์ชีวิตรวมทั้งหมดกี่คน
5) จำนวนนักเรียนจากโรงเรียน D ที่ชื่นชอบภาพยนตร์ประเภทอื่น ๆ คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนที่สำรวจทั้งหมด

บทที่ 3 | การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ



3.1 การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

3.2 การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ

3.3 ค่าวัดทางสถิติ

- 3.3.1 ค่ากลางของข้อมูล
- 3.3.2 ค่าวัดการกระจาย
- 3.3.3 ค่าวัดตำแหน่งที่ของข้อมูล

จุดมุ่งหมาย

1. สามารถวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่และแผนภาพ (ฮิสโทแกรม แผนภาพจุด แผนภาพลำต้นและใบ แผนภาพกล่อง และแผนภาพการกระจาย) พร้อมทั้งสามารถสรุปผลที่ได้จากการนำเสนอข้อมูลด้วยตารางความถี่และแผนภาพแบบต่าง ๆ
2. หาค่ากลางของข้อมูล (ค่าเฉลี่ยเลขคณิต ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก มัธยฐาน และฐานนิยม) พร้อมทั้งเลือกใช้ค่ากลางของข้อมูลที่เหมาะสมเป็นตัวแทนของข้อมูลและใช้ค่ากลางของข้อมูลในการแก้ปัญหา
3. หาค่าวัดการกระจายสัมบูรณ์ (พิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวน) และค่าวัดการกระจายสัมพัทธ์ (สัมประสิทธิ์การแปรผัน) พร้อมทั้งเลือกใช้ค่าวัดการกระจายที่เหมาะสมในการอธิบายการกระจายของข้อมูลและใช้ค่าวัดการกระจายในการแก้ปัญหา
4. หาค่าวัดตำแหน่งที่ของข้อมูล (ควอร์ไทล์และเปอร์เซ็นไทล์) พร้อมทั้งใช้ค่าวัดตำแหน่งที่ของข้อมูลในการแก้ปัญหา

บทที่ 3

การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ



ข้อมูลเชิงปริมาณเป็นข้อมูลที่พบได้มากในการดำเนินธุรกิจ การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณสามารถทำได้โดยใช้ค่าวัดทางสถิติ หรืออาจนำเสนอด้วยตารางความถี่ ฮิสโทแกรม แผนภาพจุด แผนภาพลำต้นและใบ แผนภาพกล่อง หรือแผนภาพการกระจาย เพื่อนำไปใช้ในการแปลความหมายและตัดสินใจต่อไป ตัวอย่างในธุรกิจร้านขายเครื่องดื่มและของหวาน ผู้ประกอบการต้องพิจารณาข้อมูลเกี่ยวกับปริมาณวัตถุดิบที่ใช้ในการผลิต ราคาและเวลาที่ใช้ในการสั่งซื้อวัตถุดิบแต่ละชนิด โดยอาจใช้ค่ากลางของข้อมูลที่เหมาะสมในการคาดการณ์ปริมาณวัตถุดิบแต่ละชนิดที่ต้องใช้ในแต่ละวัน สัปดาห์ หรือเดือน เพื่อที่จะดำเนินการสั่งซื้อวัตถุดิบต่าง ๆ ได้อย่างเหมาะสม ไม่เกิดเหตุการณ์ที่วัตถุดิบหมดหรือสั่งซื้อด้วยปริมาณที่มากเกินไป นอกจากนี้ ผู้ประกอบการอาจใช้ข้อมูลเกี่ยวกับปริมาณลูกค้าที่เข้ามาใช้บริการหรือซื้อสินค้า และระยะเวลาที่ลูกค้าเข้ามาใช้บริการในร้าน ในการวิเคราะห์พฤติกรรมการบริโภคของลูกค้าที่อาจเปลี่ยนแปลงไปตามแต่ละช่วงเวลาในแต่ละวัน หรือแต่ละสัปดาห์ เช่น ช่วงเวลาเร่งด่วนในตอนเช้า ช่วงพักกลางวัน และช่วงวันหยุดสุดสัปดาห์ อาจมีลูกค้าเข้ามาใช้บริการจำนวนมาก ทำให้ลูกค้าต้องใช้เวลามากขึ้นในการรอรับบริการ ผู้ประกอบการสามารถใช้ข้อมูลเหล่านี้ประกอบการวางแผนการจัดกำลังคนในแต่ละช่วงเวลา แต่ละวัน และตัดสินใจจ้างพนักงานเพิ่มได้อย่างเหมาะสม

ความรู้ก่อนหน้า

- ความหมายของสถิติศาสตร์และข้อมูล

ipst.me/10678

3.1 การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยตารางความถี่

ในบทที่ผ่านมานักเรียนได้ศึกษาเกี่ยวกับการแจกแจงความถี่ของข้อมูลเชิงคุณภาพมาแล้ว สำหรับในบทนี้นักเรียนจะได้ศึกษาการแจกแจงความถี่ของข้อมูลเชิงปริมาณซึ่งเป็นข้อมูลที่ได้จากการวัดหรือการนับค่า โดยแสดงเป็นตัวเลขหรือปริมาณที่สามารถนำไปบวก ลบ คูณ หรือหาร และเปรียบเทียบกันได้อย่างมีความหมาย ข้อมูลเชิงปริมาณสามารถใช้การแจกแจงความถี่เพื่อจัดระเบียบและนำเสนอข้อมูลได้เช่นเดียวกับข้อมูลเชิงคุณภาพ โดยการเขียนตารางความถี่สำหรับข้อมูลเชิงปริมาณมี 2 แบบ ได้แก่

1. ตารางความถี่แบบไม่ได้แบ่งข้อมูลเป็นช่วง ซึ่งเหมาะสำหรับใช้ในกรณีที่ค่าที่เป็นไปได้ของข้อมูลมีจำนวนน้อย
2. ตารางความถี่แบบแบ่งข้อมูลเป็นช่วง ซึ่งเหมาะสำหรับใช้ในกรณีที่ค่าที่เป็นไปได้ของข้อมูลมีจำนวนมาก

ตัวอย่างของตารางความถี่แบบไม่ได้แบ่งข้อมูลเป็นช่วง เช่น ในการสอบย่อยวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 10 คะแนน โดยครูให้คะแนนเป็นจำนวนเต็ม มีนักเรียนเข้าสอบ 6 คน ได้คะแนนสอบ 0, 2, 5, 5, 7 และ 10 คะแนน จะสามารถเขียนตารางความถี่สำหรับทุกค่าของคะแนนที่เป็นไปได้ซึ่งมีจำนวน 11 ค่า ได้ดังนี้

คะแนน	0	1	2	3	4	5	6	7	8	9	10
ความถี่	1	0	1	0	0	2	0	1	0	0	1

ในกรณีที่ค่าของคะแนนที่เป็นไปได้มีจำนวนมาก เช่น ในการสอบวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 100 คะแนน โดยครูให้คะแนนเป็นจำนวนเต็ม ถ้าเขียนตารางความถี่โดยใช้ทุกค่าของคะแนนที่เป็นไปได้ จะมีมากถึง 101 ค่า ซึ่งยากต่อการนำเสนอ ด้วยเหตุนี้จึงแบ่งข้อมูลที่เป็นไปได้ทั้งหมดออกเป็นช่วง ๆ และเรียกแต่ละช่วงว่า **อันตรภาคชั้น (class interval)**

ในหัวข้อนี้จะกล่าวถึงเฉพาะตารางความถี่ของข้อมูลเชิงปริมาณสำหรับข้อมูลที่มีการแบ่งเป็นอันตรภาคชั้น

ขั้นตอนการเขียนตารางความถี่ของข้อมูลเชิงปริมาณที่มีข้อมูลทั้งหมดเป็นจำนวนเต็ม

1. กำหนดจำนวนอันตรภาคชั้นเป็น k ชั้น
2. กำหนดค่าเริ่มต้นและค่าสุดท้ายที่ครอบคลุมทุกค่าของข้อมูล โดยที่ค่าเริ่มต้นคือค่าต่ำสุดหรือค่าที่น้อยกว่าค่าต่ำสุดของข้อมูล และค่าสุดท้ายคือค่าสูงสุดหรือค่าที่มากกว่าค่าสูงสุดของข้อมูล
3. คำนวณความกว้างของอันตรภาคชั้น โดยหาได้จาก

$$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}}$$

   ถ้าค่าที่คำนวณได้ไม่เป็นจำนวนเต็มให้ปัดเศษขึ้นเป็นจำนวนเต็มเสมอ
4. กำหนดอันตรภาคชั้นโดยที่
   - ชั้นแรกมีค่าเริ่มต้นที่กำหนดในข้อ 2 ถึงจำนวนที่ได้จากการนำค่าเริ่มต้นที่กำหนดในข้อ 2 บวกกับความกว้างของอันตรภาคชั้นลบด้วย 1
   - ชั้นที่สองมีค่าเริ่มต้นเป็นค่าสุดท้ายของชั้นแรกบวกด้วย 1 ถึงค่าเริ่มต้นของชั้นที่สองบวกกับความกว้างของอันตรภาคชั้นลบด้วย 1
   - ทำเช่นนี้ไปเรื่อย ๆ จนถึงชั้นที่ k (ในกรณีที่ $\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}}$ เป็นจำนวนเต็ม ค่าสุดท้ายของชั้นที่ k จะไม่เท่ากับค่าสุดท้ายที่กำหนดในข้อ 2 แต่ต้องมากกว่าหรือเท่ากับข้อมูลทุกค่า)
5. หาจำนวนข้อมูลทั้งหมดที่อยู่ในแต่ละอันตรภาคชั้น โดยทำรอยขีดแทนจำนวนไว้ในแต่ละอันตรภาคชั้น โดยปกติมักใช้รอยขีด | แทนหนึ่งค่า และเพื่อความสะดวกในการนับจำนวนข้อมูลที่อยู่ในแต่ละอันตรภาคชั้น เมื่อถึงทุก ๆ ข้อมูลที่ห้า มักนิยมทำรอยขีดแนวเฉียงหรือแนวนอนทับรอยขีดทั้งสี่ก่อนหน้านั้น ดังนี้ ~~||||~~ หรือ ~~||||~~
6. นับจำนวนข้อมูลจากรอยขีดที่ทำในข้อ 5 แล้วบันทึกจำนวนข้อมูลลงในช่องความถี่ของแต่ละอันตรภาคชั้น

ตัวอย่างที่ 1

คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 30 คน โดยคะแนนเป็นจำนวนเต็มเท่านั้น แสดงได้ดังนี้

85	112	112	123	109	85	87	123	111	112
87	126	94	92	93	91	99	121	122	128
102	118	88	106	111	98	128	94	95	129

จงเขียนตารางความถี่ของข้อมูลชุดนี้

วิธีทำ เขียนตารางความถี่ดังนี้

1. กำหนดจำนวนอันตรภาคชั้นทั้งหมด 5 ชั้น
2. กำหนดค่าเริ่มต้นเท่ากับคะแนนต่ำสุด ซึ่งคือ 85 คะแนน และค่าสุดท้ายเท่ากับคะแนนสูงสุด ซึ่งคือ 129 คะแนน
3. คำนวณความกว้างของอันตรภาคชั้น ได้ดังนี้

$$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}} = \frac{129-85}{5} = 8.8$$

ดังนั้น ความกว้างของอันตรภาคชั้นคือ 9 คะแนน

4. กำหนดอันตรภาคชั้น ได้ดังนี้

อันตรภาคชั้น	ค่าเริ่มต้น	ค่าสุดท้าย
ชั้นที่ 1	85	$85 + 9 - 1 = 93$
ชั้นที่ 2	94	$94 + 9 - 1 = 102$
ชั้นที่ 3	103	$103 + 9 - 1 = 111$
ชั้นที่ 4	112	$112 + 9 - 1 = 120$
ชั้นที่ 5	121	$121 + 9 - 1 = 129$

5. หาจำนวนข้อมูลทั้งหมดที่อยู่ในแต่ละอันตรภาคชั้นโดยทำรอยขีด ได้ดังนี้

อันตรภาคชั้น	รอยขีด
85 – 93	𝍸 |||
94 – 102	𝍸 |
103 – 111	||||
112 – 120	||||
121 – 129	𝍸 |||

6. นับจำนวนข้อมูลจากรอยขีดที่ทำในข้อ 5 จะได้ตารางความถี่ ดังนี้

อันตรภาคชั้น	ความถี่
85 – 93	8
94 – 102	6
103 – 111	4
112 – 120	4
121 – 129	8

■

จากตารางความถี่ในตัวอย่างที่ 1 สามารถหาความถี่สะสมในแต่ละอันตรภาคชั้น ซึ่งคือผลรวมของความถี่ของอันตรภาคชั้นนั้นกับความถี่ของอันตรภาคชั้นก่อนหน้าทั้งหมด ได้ดังนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม
85 – 93	8	8
94 – 102	6	14
103 – 111	4	18
112 – 120	4	22
121 – 129	8	30

นอกจากนี้ สามารถหาความถี่สัมพัทธ์และความถี่สะสมสัมพัทธ์ในแต่ละอันตรภาคชั้น ซึ่งความถี่สะสมสัมพัทธ์ในแต่ละอันตรภาคชั้นคือผลรวมของความถี่สัมพัทธ์ของอันตรภาคชั้นนั้นกับความถี่สัมพัทธ์ของอันตรภาคชั้นก่อนหน้าทั้งหมด ได้ดังนี้

อันตรภาคชั้น	ความถี่	ความถี่สะสม	ความถี่สัมพัทธ์		ความถี่สะสมสัมพัทธ์	
			สัดส่วน	ร้อยละ	สัดส่วน	ร้อยละ
85 – 93	8	8	$\frac{8}{30} \approx 0.27$	27	0.27	27
94 – 102	6	14	$\frac{6}{30} = 0.20$	20	0.47	47
103 – 111	4	18	$\frac{4}{30} \approx 0.13$	13	0.60	60
112 – 120	4	22	$\frac{4}{30} \approx 0.13$	13	0.73	73
121 – 129	8	30	$\frac{8}{30} \approx 0.27$	27	1	100

จากตาราง อาจสรุปได้ว่า

- นักเรียนที่ได้คะแนนตั้งแต่ 85 ถึง 93 คะแนน มี 8 คน คิดเป็นประมาณร้อยละ 27 ของจำนวนนักเรียน 30 คนนี้
- นักเรียนที่ได้คะแนนตั้งแต่ 94 ถึง 102 คะแนน มี 6 คน คิดเป็นร้อยละ 20 ของจำนวนนักเรียน 30 คนนี้
- นักเรียนที่ได้คะแนนน้อยกว่าหรือเท่ากับ 102 คะแนน มี 14 คน คิดเป็นประมาณร้อยละ 47 ของจำนวนนักเรียน 30 คนนี้
- นักเรียนที่ได้คะแนนตั้งแต่ 112 ถึง 120 คะแนน มี 4 คน คิดเป็นประมาณร้อยละ 13 ของจำนวนนักเรียน 30 คนนี้
- นักเรียนที่ได้คะแนนน้อยกว่าหรือเท่ากับ 120 คะแนน มี 22 คน คิดเป็นประมาณร้อยละ 73 ของจำนวนนักเรียน 30 คนนี้

หมายเหตุ จากตารางความถี่ ถ้าต้องการหาความกว้างของแต่ละอันตรภาคชั้น สามารถหาได้จากค่าสุดท้ายในแต่ละอันตรภาคชั้นลบด้วยค่าเริ่มต้นในอันตรภาคชั้นนั้นบวกด้วย 1

ในการกำหนดจำนวนอันตรภาคชั้นและความกว้างของอันตรภาคชั้นมีข้อสังเกต ดังนี้

1. ถ้าข้อมูลบางค่าแตกต่างไปจากค่าอื่น ๆ ในข้อมูลชุดนั้นมาก เช่น ถ้ามีผู้เข้าสอบคนหนึ่งสอบได้ 5 คะแนน ในขณะที่คนอื่น ๆ ได้มากกว่าหรือเท่ากับ 40 คะแนน ควรกำหนดอันตรภาคชั้นแรกเป็น **อันตรภาคชั้นเปิด (open-ended class interval)** เช่น ในกรณีนี้อาจกำหนดอันตรภาคชั้นแรกเป็น “น้อยกว่า 40” ดังนี้

อันตรภาคชั้น (คะแนน)	ความถี่
น้อยกว่า 40	1
40 – 49	2
50 – 59	6
60 – 69	20
70 – 79	21
80 – 89	8
90 – 99	2

2. ความกว้างของแต่ละอันตรภาคชั้นไม่จำเป็นต้องเท่ากันทั้งหมด ทั้งนี้ ขึ้นอยู่กับวัตถุประสงค์ของการใช้ข้อมูลและอาจขึ้นอยู่กับความแตกต่างระหว่างข้อมูลชุดนั้น เช่น โรงภาพยนตร์แห่งหนึ่งต้องการวิเคราะห์พฤติกรรมการชมภาพยนตร์ของผู้ใช้บริการแต่ละกลุ่มอายุ เพื่อวางแผนการตลาดให้สอดคล้องกับผู้ใช้บริการแต่ละกลุ่ม จึงเขียนแสดงผลสำรวจจำนวนผู้ใช้บริการ จำแนกตามอายุ ดังนี้

อายุ (ปี)	จำนวนผู้ใช้บริการ (คน)
น้อยกว่า 12	50
12 – 17	650
18 – 24	12,560
25 – 34	8,720
35 – 49	5,838
50 – 60	2,554
มากกว่า 60	110

ในกรณีที่ข้อมูลไม่เป็นจำนวนเต็ม การกำหนดอันตรภาคชั้นอาจกำหนดให้อยู่ในรูปช่วง ดังแสดงในตัวอย่างต่อไปนี้

ตัวอย่างที่ 2

จากการสำรวจจำนวนชั่วโมงเฉลี่ยของการทำงานใน 1 สัปดาห์ ของคนงานในโรงงาน 8 แห่ง ในแต่ละเขตอุตสาหกรรม ได้ข้อมูลดังนี้

จำนวนชั่วโมงเฉลี่ยของการทำงานต่อคนต่อสัปดาห์ในแต่ละโรงงาน					
เขต ก	ชั่วโมง	เขต ข	ชั่วโมง	เขต ค	ชั่วโมง
คนงาน ก1	35.0	คนงาน ข1	40.0	คนงาน ค1	43.0
คนงาน ก2	48.0	คนงาน ข2	50.0	คนงาน ค2	48.8
คนงาน ก3	45.0	คนงาน ข3	35.4	คนงาน ค3	43.3
คนงาน ก4	43.0	คนงาน ข4	38.8	คนงาน ค4	53.1
คนงาน ก5	38.2	คนงาน ข5	40.2	คนงาน ค5	35.6
คนงาน ก6	50.0	คนงาน ข6	45.0	คนงาน ค6	41.1
คนงาน ก7	39.8	คนงาน ข7	45.0	คนงาน ค7	34.8
คนงาน ก8	40.7	คนงาน ข8	40.0	คนงาน ค8	51.0

จงเขียนตารางความถี่ของข้อมูลชุดนี้

วิธีทำ เขียนตารางความถี่ดังนี้

1. กำหนดจำนวนอันตรภาคชั้นทั้งหมด 5 ชั้น
2. เนื่องจากข้อมูลมีค่าต่ำสุดคือ 34.8 ชั่วโมงต่อสัปดาห์ และค่าสูงสุดคือ 53.1 ชั่วโมงต่อสัปดาห์ จึงกำหนดค่าเริ่มต้นคือ 30 ชั่วโมงต่อสัปดาห์ และค่าสุดท้ายคือ 55 ชั่วโมงต่อสัปดาห์
3. คำนวณความกว้างของอันตรภาคชั้น ได้ดังนี้

   $$\frac{\text{ค่าสุดท้าย} - \text{ค่าเริ่มต้น}}{\text{จำนวนอันตรภาคชั้น}} = \frac{55-30}{5} = 5$$

   ดังนั้น ความกว้างของอันตรภาคชั้นคือ 5 ชั่วโมงต่อสัปดาห์
4. กำหนดอันตรภาคชั้นในรูปช่วง โดยแบ่งเป็น 5 ชั้น พร้อมทั้งหาจำนวนข้อมูลทั้งหมดที่อยู่ในแต่ละอันตรภาคชั้นโดยทำรอยขีด ได้ดังนี้

จำนวนชั่วโมง (x)	รอยขีด
$30 \le x < 35$	𝍷
$35 \le x < 40$	𝍸 𝍷
$40 \le x < 45$	𝍸 𝍷𝍷𝍷
$45 \le x < 50$	𝍸
$50 \le x < 55$	𝍷𝍷𝍷𝍷

5. นับจำนวนข้อมูลจากรอยขีดที่ทำในข้อ 4 จะได้ตารางความถี่ ดังนี้

จำนวนชั่วโมง (x)	ความถี่
$30 \le x < 35$	1
$35 \le x < 40$	6
$40 \le x < 45$	8
$45 \le x < 50$	5
$50 \le x < 55$	4

■

การกำหนดจำนวนอันตรภาคชั้นในการเขียนตารางความถี่ไม่มีกฎเกณฑ์แน่นอนตายตัวขึ้นอยู่กับลักษณะการกระจายของข้อมูลหรือความแตกต่างระหว่างข้อมูลว่ามีมากน้อยเพียงใด และอาจขึ้นอยู่กับรายละเอียดของข้อมูลที่ต้องการทราบด้วย เช่น ถ้าข้อมูลมีความแตกต่างกันมาก มักจะกำหนดจำนวนอันตรภาคชั้นให้น้อย เพื่อไม่ให้อันตรภาคชั้นส่วนใหญ่มีความถี่เป็นศูนย์ เนื่องจากไม่มีข้อมูลใดอยู่ในอันตรภาคชั้นนั้นเลย หรือถ้าต้องการทราบรายละเอียดเกี่ยวกับข้อมูลนั้นมาก ก็ควรกำหนดจำนวนอันตรภาคชั้นให้มากขึ้น โดยทั่วไปจำนวนอันตรภาคชั้นที่นิยมใช้จะอยู่ระหว่าง 7 ถึง 15 ชั้น หรือส่วนใหญ่จะไม่ต่ำกว่า 5 ชั้น

หมายเหตุ สำหรับข้อมูลที่ไม่เป็นจำนวนเต็มอาจกำหนดช่วงของอันตรภาคชั้นในรูป $a < x \le b$ แต่ในที่นี้จะเขียนในรูป $a \le x < b$ เท่านั้น

แบบฝึกหัด 3.1

1. จำนวนเซลล์เม็ดเลือดขาว (เซลล์ต่อเลือด 1 ลูกบาศก์มิลลิเมตร) ของครูระดับมัธยมศึกษาตอนปลายในภาคตะวันออกเฉียงเหนือจำนวน 40 คน แสดงได้ดังนี้

9,328	8,100	9,517	8,909	9,161
9,015	8,848	9,612	8,182	9,107
9,006	9,207	9,418	8,708	9,247
8,594	8,820	8,401	8,409	8,701
8,767	8,969	9,297	8,985	8,735
9,315	9,419	9,123	8,716	8,501
8,974	8,412	9,378	8,112	9,432
9,211	9,078	8,988	8,811	8,841

จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 9 ชั้น ค่าเริ่มต้นและค่าสุดท้ายเท่ากับ 8,000 และ 9,800 เซลล์ต่อเลือด 1 ลูกบาศก์มิลลิเมตร ตามลำดับ พร้อมทั้งสรุปผลที่ได้

เสริมสมอง : เซลล์เม็ดเลือดขาว

เซลล์เม็ดเลือดขาวทำหน้าที่ป้องกันและทำลายเชื้อโรคหรือสิ่งแปลกปลอม โดยปกติแล้วในเลือด 1 ลูกบาศก์มิลลิเมตร จะมีเซลล์เม็ดเลือดขาวอยู่ประมาณ 5,000 – 10,000 เซลล์

2. จำนวนอีเมลที่กนกวรรณได้รับในแต่ละวันตั้งแต่วันที่ 1 – 31 กรกฎาคม พ.ศ. 2561 แสดงได้ดังนี้

วันที่	จำนวนอีเมลที่ได้รับ (ฉบับ)
1	28
2	18
3	15
4	23
5	26
6	17
7	9
8	27
9	16
10	24
11	23
12	24
13	11
14	43
15	20
16	17

วันที่	จำนวนอีเมลที่ได้รับ (ฉบับ)
17	17
18	25
19	18
20	32
21	18
22	15
23	20
24	22
25	18
26	30
27	17
28	13
29	25
30	26
31	20

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 6 ชั้น ค่าเริ่มต้นเท่ากับ 8 ฉบับ และค่าสุดท้ายเท่ากับ 44 ฉบับ
2) อันตรภาคชั้นใดมีความถี่สูงที่สุด
3) อันตรภาคชั้นที่ 5 มีความถี่สัมพัทธ์ในรูปร้อยละเป็นเท่าใด
4) จำนวนวันที่กนกวรรณได้รับอีเมลน้อยกว่า 32 ฉบับ คิดเป็นร้อยละเท่าใดของจำนวนวันทั้งหมดในเดือนกรกฎาคม พ.ศ. 2561

3. ซูเปอร์มาร์เก็ตของห้างสรรพสินค้าแห่งหนึ่งเก็บข้อมูลยอดชำระเงิน (บาท) ของลูกค้าจำนวน 50 คน โดยเรียงข้อมูลจากน้อยไปมากได้ดังนี้

150	189	1,112	1,137	1,156	1,182	1,201	1,209	1,217	1,221
1,236	1,248	1,256	1,261	1,262	1,287	1,291	1,301	1,316	1,323
1,336	1,341	1,347	1,352	1,352	1,360	1,369	1,372	1,387	1,391
1,402	1,409	1,410	1,417	1,421	1,423	1,430	1,439	1,446	1,446
1,454	1,456	1,460	1,468	1,508	1,524	1,546	1,581	1,592	1,672

1) จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สัมพัทธ์ของข้อมูลชุดนี้ โดยกำหนดอันตรภาคชั้นเป็น ต่ำกว่า 1100, 1100 – 1199, 1200 – 1299, 1300 – 1399, 1400 – 1499, 1500 – 1599 และ 1600 – 1699
2) ลูกค้ามียอดชำระเงินอยู่ในอันตรภาคชั้นใดมากที่สุด และคิดเป็นร้อยละเท่าใดของจำนวนลูกค้าที่เก็บข้อมูลทั้งหมด
3) จำนวนลูกค้าที่มียอดชำระเงินต่ำกว่า 1,200 บาท มากกว่า น้อยกว่า หรือเท่ากับจำนวนลูกค้าที่มียอดชำระเงินตั้งแต่ 1,500 บาทขึ้นไป
4) ผู้จัดการซูเปอร์มาร์เก็ตแห่งนี้จะสรุปเกี่ยวกับยอดชำระเงินของลูกค้าได้อย่างไรบ้าง

4. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวนทั้งหมด 60 คน ซึ่งมีคะแนนเต็ม 100 คะแนน แสดงได้ดังนี้

84	79	65	78	78	62	80	67	82	73
81	68	60	74	67	75	92	80	71	62
52	62	76	76	65	63	68	51	48	53
71	75	74	77	68	73	61	66	96	79
67	70	61	81	57	62	57	69	60	76
81	93	75	72	60	65	56	75	88	35

1) จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ของข้อมูลชุดนี้ โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 7 ชั้น ค่าเริ่มต้นเท่ากับ 30 คะแนน และค่าสุดท้ายเท่ากับ 101 คะแนน
2) มีนักเรียนที่ได้คะแนนตั้งแต่ 85 คะแนนขึ้นไปกี่คน
3) นักเรียนที่ได้คะแนนน้อยกว่า 52 คะแนน คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด
4) นักเรียนที่ได้คะแนนตั้งแต่ 52 ถึง 84 คะแนน คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด (ตอบเป็นจำนวนเต็ม)

5. จำนวนประชากรที่อาศัยอยู่ในหมู่บ้านต่าง ๆ ของอำเภอหนึ่ง จำนวน 66 หมู่บ้าน แสดงได้ดังนี้

345	494	604	730	468	607	395	494	530	563
515	505	468	461	560	490	556	444	523	549
624	439	384	402	354	629	545	472	582	596
585	523	593	439	505	475	574	417	629	420
574	486	527	611	578	490	490	585	516	505
431	541	523	523	468	549	479	604	464	691
575	585	603	417	446	406				

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดให้ 341 – 380 เป็นอันตรภาคชั้นแรก และแต่ละอันตรภาคชั้นมีความกว้างเท่ากัน
2) มีหมู่บ้านที่มีประชากรอาศัยอยู่ต่ำกว่า 501 คน กี่หมู่บ้าน
3) มีหมู่บ้านที่มีประชากรอาศัยอยู่ตั้งแต่ 381 ถึง 580 คน กี่หมู่บ้าน
4) จำนวนหมู่บ้านที่มีประชากรอาศัยอยู่มากกว่า 660 คน คิดเป็นร้อยละเท่าใดของจำนวนหมู่บ้านทั้งหมด

3.2 การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณด้วยแผนภาพ

ในการวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ นอกจากจะใช้ตารางความถี่แล้ว ยังสามารถใช้แผนภาพได้อีกด้วย โดยการใช้แผนภาพแสดงการแจกแจงความถี่ของข้อมูล จะทำให้เห็นการแจกแจงของข้อมูลได้ชัดเจนมากกว่าการพิจารณาจากตารางความถี่

ฮิสโทแกรม

ฮิสโทแกรม (histogram) เป็นการนำเสนอข้อมูลเชิงปริมาณที่สร้างจากตารางความถี่ โดยใช้แท่งสี่เหลี่ยมมุมฉากที่เรียงติดกันบนแกนนอน เมื่อแกนนอนแทนค่าของข้อมูล ความสูงของแท่งสี่เหลี่ยมมุมฉากจะแสดงความถี่ของข้อมูล ซึ่งการแสดงความถี่ของข้อมูลอาจนำเสนอความถี่ของข้อมูลเพียงค่าเดียวหรือข้อมูลในแต่ละอันตรภาคชั้น โดยความกว้างของแท่งสี่เหลี่ยมมุมฉากแต่ละแท่งจะสอดคล้องกับความกว้างของแต่ละอันตรภาคชั้นของตารางความถี่

พิจารณาสถานการณ์ต่อไปนี้

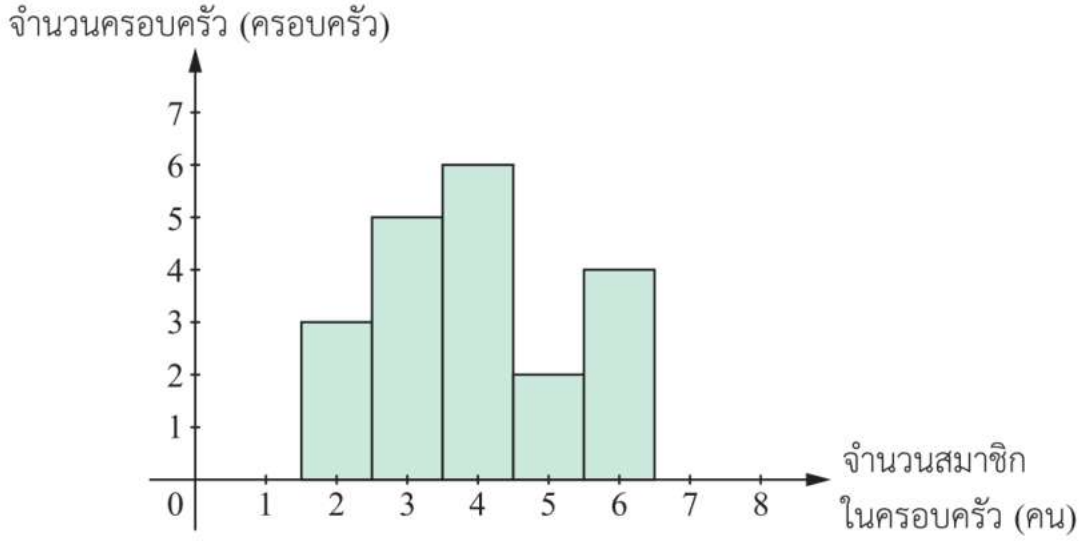
คณะกรรมการหมู่บ้านจัดสรรแห่งหนึ่งสำรวจข้อมูลจำนวนสมาชิกในแต่ละครอบครัว จำนวน 20 ครอบครัว โดยเรียงข้อมูลจากน้อยไปมากได้ดังนี้

2	2	2	3	3	3	3	3	4	4
4	4	4	4	5	5	6	6	6	6

จากข้อมูลข้างต้น เขียนตารางความถี่ได้ดังนี้

จำนวนสมาชิกในครอบครัว (คน)	จำนวนครอบครัว (ครอบครัว)
2	3
3	5
4	6
5	2
6	4
รวม	**20**

สามารถนำเสนอข้อมูลข้างต้นโดยใช้ฮิสโทแกรมได้ดังนี้



จากฮิสโทแกรม จะเห็นว่าครอบครัวที่มีสมาชิก 4 คน มีจำนวนมากที่สุด ไม่มีครอบครัวใดเลยที่มีสมาชิกมากกว่า 6 คน และไม่มีครอบครัวใดเลยที่มีสมาชิกเพียง 1 คน

ข้อสังเกต

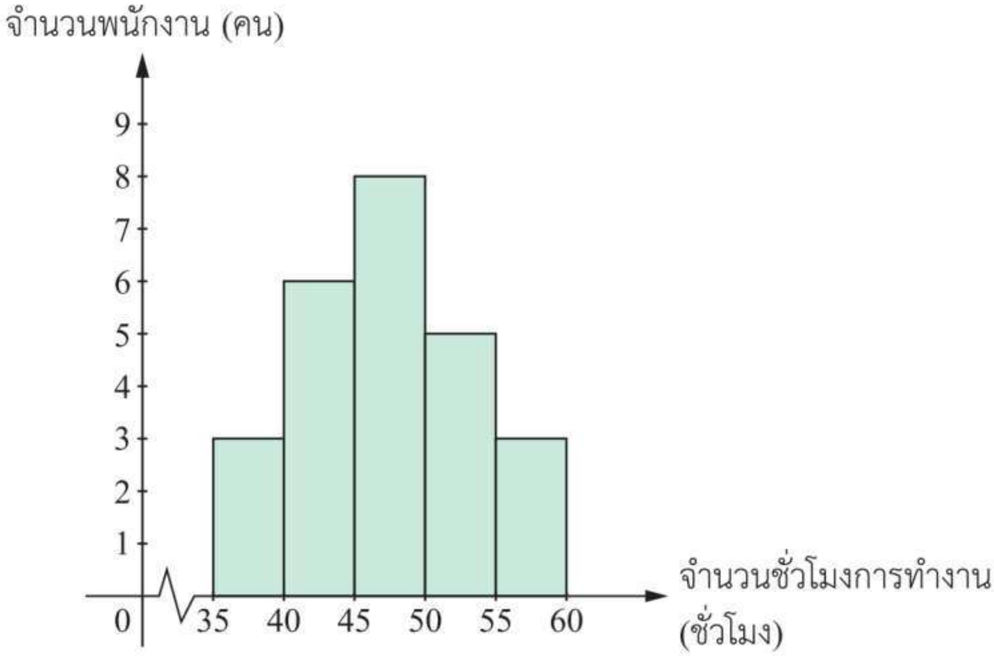
1. จุดกึ่งกลางของฐานของแต่ละแท่งสี่เหลี่ยมมุมฉากแทนข้อมูลแต่ละค่า โดยในที่นี้คือจำนวนสมาชิกในครอบครัว ซึ่งได้แก่ 2, 3, 4, 5 และ 6 คน
2. ความสูงของแต่ละแท่งสี่เหลี่ยมมุมฉากแทนความถี่ของข้อมูลแต่ละค่า โดยในที่นี้คือจำนวนครอบครัว ซึ่งได้แก่ 3, 5, 6, 2 และ 4 ครอบครัว
3. ฮิสโทแกรมมีลักษณะคล้ายแผนภูมิแท่ง แต่แผนภูมิแท่งใช้นำเสนอข้อมูลเชิงคุณภาพ

นอกจากนี้สามารถเขียนฮิสโทแกรมจากตารางความถี่ที่มีการแบ่งข้อมูลเป็นอันตรภาคชั้นได้

พิจารณาสถานการณ์ต่อไปนี้
ฝ่ายทรัพยากรบุคคลของหน่วยงานแห่งหนึ่งได้เก็บข้อมูลจำนวนชั่วโมงการทำงานในหนึ่งสัปดาห์ของพนักงานจำนวน 25 คน ได้ข้อมูลดังตารางความถี่ต่อไปนี้

จำนวนชั่วโมงการทำงาน (x)	จำนวนพนักงาน (คน)
$35 \le x < 40$	3
$40 \le x < 45$	6
$45 \le x < 50$	8
$50 \le x < 55$	5
$55 \le x < 60$	3
รวม	**25**

สามารถนำเสนอข้อมูลข้างต้นโดยใช้ฮิสโทแกรมได้ดังนี้



จากฮิสโทแกรม จะเห็นว่าในหนึ่งสัปดาห์ พนักงานที่มีชั่วโมงการทำงานตั้งแต่ 45 ชั่วโมง แต่น้อยกว่า 50 ชั่วโมง มีจำนวนมากที่สุด และจำนวนพนักงานที่มีชั่วโมงการทำงานตั้งแต่ 35 ชั่วโมง แต่น้อยกว่า 40 ชั่วโมง เท่ากับจำนวนพนักงานที่มีชั่วโมงการทำงานตั้งแต่ 55 ชั่วโมง แต่น้อยกว่า 60 ชั่วโมง

ข้อสังเกต เนื่องจากการนำเสนอข้อมูลด้วยตารางความถี่ที่มีการแบ่งข้อมูลเป็นอันตรภาคชั้น จะไม่สามารถบอกได้ว่าข้อมูลที่เก็บรวบรวมมามีค่าใดบ้าง และแต่ละค่ามีความถี่เท่าใด เช่น จากตารางความถี่ข้างต้น มีพนักงาน 3 คน ที่มีชั่วโมงการทำงานตั้งแต่ 35 ชั่วโมง แต่น้อยกว่า 40 ชั่วโมง แต่ไม่สามารถทราบได้ว่าพนักงาน 3 คนนี้ แต่ละคนทำงานกี่ชั่วโมง ดังนั้น ฮิสโทแกรมที่สร้างจากตารางความถี่ที่มีการแบ่งข้อมูลเป็นอันตรภาคชั้นก็จะไม่สามารถบอกรายละเอียดของข้อมูลแต่ละค่าได้เช่นเดียวกัน

ฮิสโทแกรมที่เขียนจากสถานการณ์ข้างต้นเป็นการนำเสนอข้อมูลที่ไม่เป็นจำนวนเต็ม เช่น จำนวนชั่วโมงการทำงานของพนักงานคนหนึ่งอาจเป็น 37.42 ชั่วโมง แต่ในกรณีที่ข้อมูลเป็นจำนวนเต็ม และตารางความถี่มีการแบ่งข้อมูลเป็นอันตรภาคชั้น เช่น ข้อมูลคะแนนสอบในตัวอย่างที่ 1 การเขียนฮิสโทแกรมเพื่อให้แท่งสี่เหลี่ยมมุมฉากแต่ละรูปเรียงติดกัน จะต้องหาขอบล่างของชั้นและขอบบนของชั้นของแต่ละอันตรภาคชั้นก่อน โดยที่

ขอบล่างของชั้น (lower class boundary) คือ ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุดในชั้นก่อนหน้ากับค่าของข้อมูลที่น้อยที่สุดในชั้นนั้น

ขอบบนของชั้น (upper class boundary) คือ ค่ากึ่งกลางระหว่างค่าของข้อมูลที่มากที่สุดในชั้นนั้นกับค่าของข้อมูลที่น้อยที่สุดในชั้นถัดไป

จากนั้น สามารถเขียนฮิสโทแกรมได้ โดยให้ความกว้างของแท่งสี่เหลี่ยมมุมฉากแต่ละแท่งเท่ากับผลต่างของขอบบนของชั้นและขอบล่างของชั้นของแต่ละอันตรภาคชั้น

หมายเหตุ การหาขอบล่างของชั้นของอันตรภาคชั้นแรก ให้พิจารณาเสมือนว่ามีอันตรภาคชั้นที่ต่ำกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น ในทำนองเดียวกัน การหาขอบบนของชั้นในอันตรภาคชั้นสุดท้าย ให้พิจารณาเสมือนว่ามีอันตรภาคชั้นที่สูงกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น

จากตารางความถี่ในตัวอย่างที่ 1 สามารถแสดงวิธีการหาขอบล่างของชั้นและขอบบนของชั้นของแต่ละอันตรภาคชั้น ได้ดังนี้

พิจารณาอันตรภาคชั้น 85 – 93

- สามารถหาขอบล่างของชั้นโดยพิจารณาเสมือนว่ามีอันตรภาคชั้นที่ต่ำกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น ซึ่งคือ 76 – 84 จะได้ ขอบล่างของชั้นคือ $\frac{84+85}{2} = 84.5$
- เนื่องจากอันตรภาคชั้นถัดไปคือ 94 – 102 จะได้ ขอบบนของชั้นคือ $\frac{93+94}{2} = 93.5$

พิจารณาอันตรภาคชั้น 94 – 102

- เนื่องจากอันตรภาคชั้นก่อนหน้าคือ 85 – 93 จะได้ ขอบล่างของชั้นคือ $\frac{93+94}{2} = 93.5$
- เนื่องจากอันตรภาคชั้นถัดไปคือ 103 – 111 จะได้ ขอบบนของชั้นคือ $\frac{102+103}{2} = 102.5$

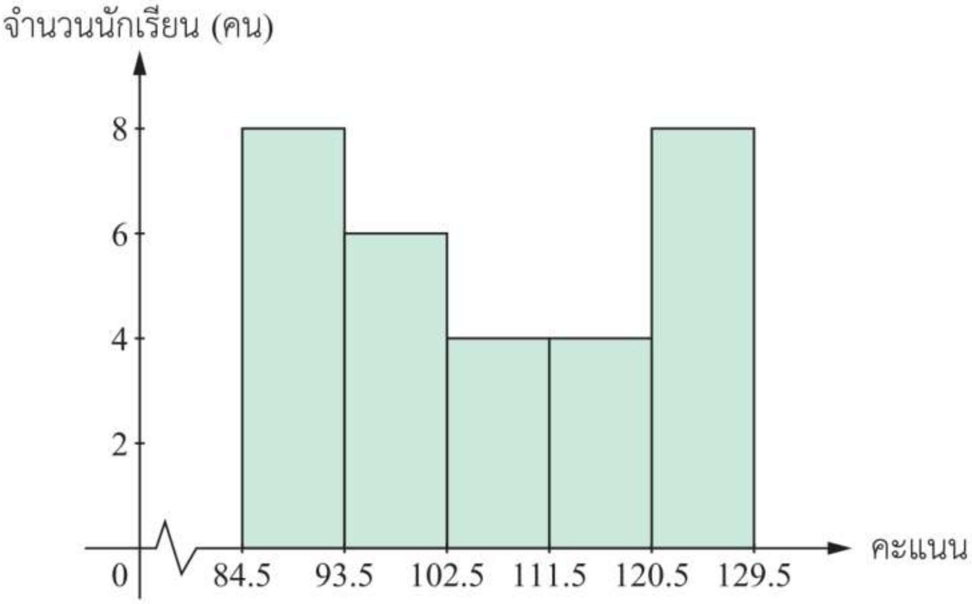
ทำเช่นนี้ไปเรื่อย ๆ จนถึงอันตรภาคชั้นสุดท้าย ซึ่งคือ 121 – 129

- เนื่องจากอันตรภาคชั้นก่อนหน้าคือ 112 – 120 จะได้ ขอบล่างของชั้นคือ $\frac{120+121}{2} = 120.5$
- สามารถหาขอบบนของชั้นโดยพิจารณาเสมือนว่ามีอันตรภาคชั้นที่สูงกว่าอันตรภาคชั้นนี้อีกหนึ่งชั้น ซึ่งคือ 130 – 138 จะได้ ขอบบนของชั้นคือ $\frac{129+130}{2} = 129.5$

จะได้ตารางความถี่ที่แสดงขอบล่างของชั้นและขอบบนของชั้นของแต่ละอันตรภาคชั้น ดังนี้

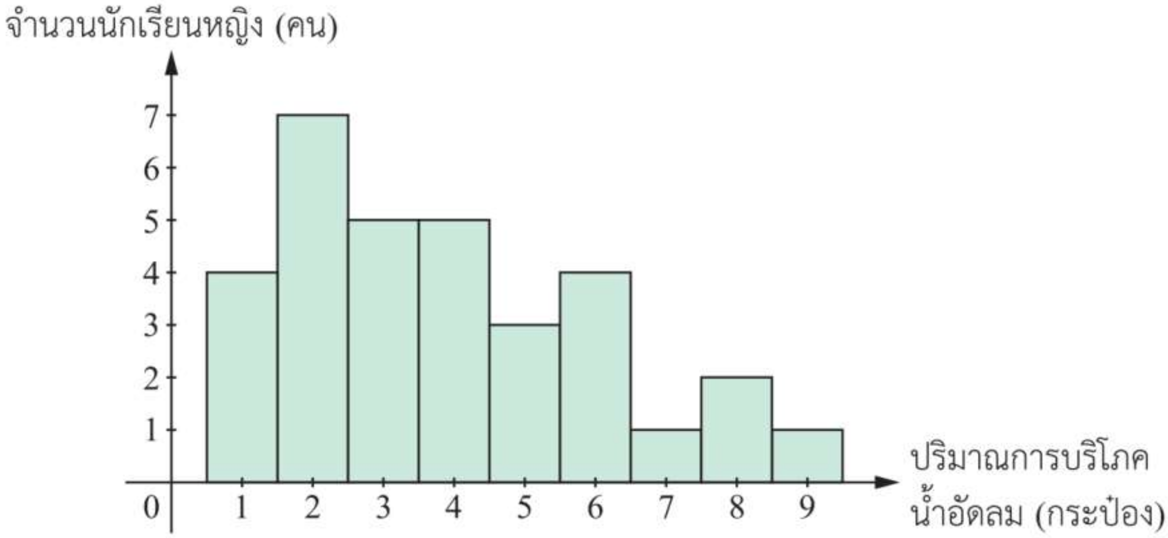
คะแนน	ขอบล่าง – ขอบบน	จำนวนนักเรียน (คน)
85 – 93	84.5 – 93.5	8
94 – 102	93.5 – 102.5	6
103 – 111	102.5 – 111.5	4
112 – 120	111.5 – 120.5	4
121 – 129	120.5 – 129.5	8

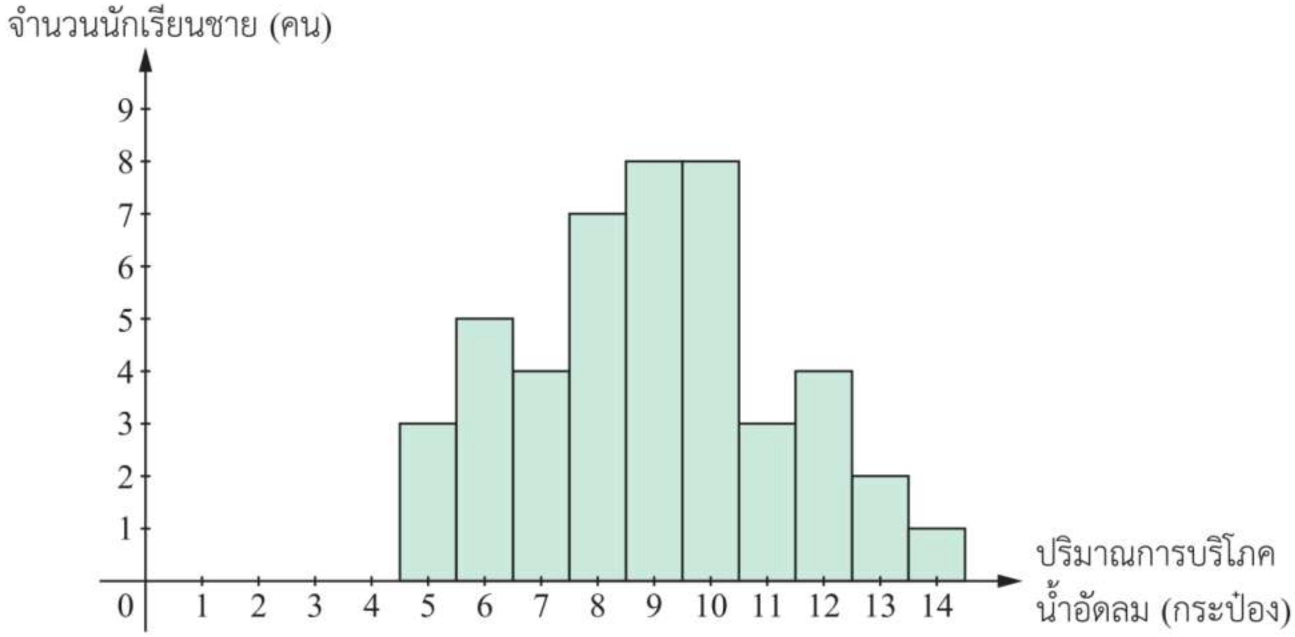
และสามารถเขียนฮิสโทแกรม ได้ดังนี้



ตัวอย่างที่ 3

ปริมาณการบริโภคน้ำอัดลม (กระป๋อง) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในโรงเรียนแห่งหนึ่งในหนึ่งสัปดาห์ แสดงด้วยฮิสโทแกรมได้ดังนี้





1) นักเรียนหญิงจำนวนมากที่สุดบริโภคน้ำอัดลมกี่กระป๋องในหนึ่งสัปดาห์
2) มีนักเรียนชายบริโภคน้ำอัดลมมากที่สุดกี่กระป๋องในหนึ่งสัปดาห์
3) จงสรุปลักษณะที่น่าสนใจเกี่ยวกับพฤติกรรมการบริโภคน้ำอัดลมของนักเรียนหญิงและนักเรียนชาย ในโรงเรียนแห่งนี้

วิธีทำ 1) จากฮิสโทแกรมแสดงปริมาณการบริโภคน้ำอัดลมของนักเรียนหญิงในโรงเรียนแห่งนี้ จะได้ว่านักเรียนหญิงบริโภคน้ำอัดลม 2 กระป๋องในหนึ่งสัปดาห์มีจำนวนมากที่สุด

2) จากฮิสโทแกรมแสดงปริมาณการบริโภคน้ำอัดลมของนักเรียนชายในโรงเรียนแห่งนี้ จะได้ว่ามีนักเรียนชายบริโภคน้ำอัดลมมากที่สุด 14 กระป๋องในหนึ่งสัปดาห์

3) จากฮิสโทแกรม สรุปได้ว่าในหนึ่งสัปดาห์ นักเรียนหญิงส่วนใหญ่บริโภคน้ำอัดลมน้อยกว่า 6 กระป๋อง ในขณะที่นักเรียนชายส่วนใหญ่บริโภคน้ำอัดลมมากกว่า 6 กระป๋อง ■

แผนภาพจุด

แผนภาพจุด (dot plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้จุดหรือวงกลมเล็ก ๆ แทนข้อมูลแต่ละตัว เขียนเรียงไว้เหนือเส้นในแนวนอนที่มีสเกล จุดหรือวงกลมเล็ก ๆ ดังกล่าวจะเรียงกันในแนวตั้งตรงกับตำแหน่งซึ่งแสดงค่าของข้อมูลแต่ละตัว

พิจารณาสถานการณ์ต่อไปนี้

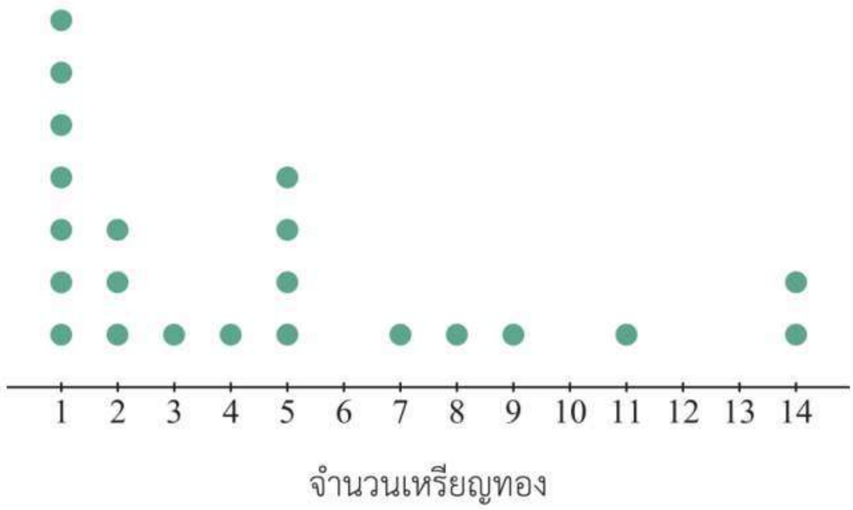
จากการแข่งขันกีฬาโอลิมปิกฤดูหนาว 2018 ซึ่งจัดขึ้นที่สาธารณรัฐเกาหลี ใน พ.ศ. 2561 สามารถแสดงข้อมูลจำนวนเหรียญทองของประเทศที่ได้เหรียญทองจำนวน 22 ประเทศ โดยเรียงข้อมูลจากมากไปน้อยได้ดังนี้

14	14	11	9	8	7	5	5	5	5	4
3	2	2	2	1	1	1	1	1	1	1

จากข้อมูลข้างต้น เขียนตารางความถี่ได้ดังนี้

จำนวนเหรียญทอง (เหรียญ)	จำนวนประเทศ
1	7
2	3
3	1
4	1
5	4
7	1
8	1
9	1
11	1
14	2

จากตารางความถี่ข้างต้น สามารถเขียนแผนภาพจุดได้ดังนี้



จากแผนภาพจุด จะเห็นว่าประเทศที่ได้เหรียญทอง 1 เหรียญ มีจำนวนมากที่สุด และมีเพียง 3 ประเทศที่ได้เหรียญทองมากกว่า 10 เหรียญ

ข้อสังเกต

1. ค่าที่เป็นไปได้ทั้งหมดของข้อมูลจะปรากฏบนแกนนอนของแผนภาพจุด ถึงแม้ว่าบางค่าอาจมีความถี่เป็นศูนย์ เช่น จากสถานการณ์ข้างต้น จะเห็นว่าไม่มีประเทศใดเลยที่ได้เหรียญทอง 6, 10, 12 และ 13 เหรียญ แต่ค่าดังกล่าวยังคงปรากฏบนแกนนอนของแผนภาพจุด
2. ในแผนภาพจุดอาจกำหนดให้ค่าเริ่มต้นเป็นค่าต่ำสุดและค่าสุดท้ายเป็นค่าสูงสุดของข้อมูลได้

เสริมสมอง : โอลิมปิกฤดูหนาว

โอลิมปิกฤดูหนาว (Olympic Winter Games) เป็นการแข่งขันกีฬาฤดูหนาวที่จัดขึ้นทุก 4 ปี โดยจัดห่างจากโอลิมปิกฤดูร้อน 2 ปี ในความควบคุมของคณะกรรมการโอลิมปิกสากล (International Olympic Committee: IOC) ประเภทกีฬาที่แข่งขันในโอลิมปิกฤดูหนาว เช่น สกีอัลไพน์ (alpine skiing) สโนว์บอร์ด (snowboard) ฮอกกี้น้ำแข็ง (ice hockey) ฟิกเกอร์สเกต (figure skating) จำนวนประเทศที่เข้าร่วมโอลิมปิกฤดูหนาวน้อยกว่าจำนวนประเทศที่เข้าร่วมโอลิมปิกฤดูร้อนมาก เนื่องจากประเทศส่วนใหญ่มีสภาพภูมิอากาศไม่เอื้ออำนวยต่อกีฬาฤดูหนาว

ตัวอย่างที่ 4

จำนวนชั่วโมงการออกกำลังกายของครูในโรงเรียนแห่งหนึ่งในหนึ่งสัปดาห์ แสดงด้วยแผนภาพจุดได้ดังนี้



จงสรุปลักษณะที่น่าสนใจเกี่ยวกับจำนวนชั่วโมงการออกกำลังกายของครูในโรงเรียนแห่งนี้ในหนึ่งสัปดาห์

วิธีทำ จากแผนภาพจุด สรุปได้ว่า ในหนึ่งสัปดาห์ครูส่วนใหญ่ออกกำลังกายน้อยกว่า 4 ชั่วโมง และมีครูจำนวน 2 คน ที่ออกกำลังมากถึง 12 ชั่วโมงต่อสัปดาห์ ■

แผนภาพลำต้นและใบ

แผนภาพลำต้นและใบ (stem and leaf plot) เป็นการนำเสนอข้อมูลเชิงปริมาณโดยใช้แผนภาพที่มีการแสดงข้อมูลโดยเรียงลำดับจากน้อยไปมากและแบ่งการแสดงข้อมูลออกเป็นสองส่วนที่เรียกว่า ส่วน**ลำต้น** และส่วน**ใบ** ในที่นี้กำหนดส่วนใบเป็นเลขโดดในหลักหน่วย และตัวเลขที่เหลือเป็นส่วนลำต้น เช่น 298 จะมี 29 เป็นส่วนลำต้น และ 8 เป็นส่วนใบ

หมายเหตุ ในที่นี้จะใช้แผนภาพลำต้นและใบนำเสนอข้อมูลเชิงปริมาณที่เป็นจำนวนเต็มบวกเท่านั้น

พิจารณาสถานการณ์ต่อไปนี้

ร้านอาหารแห่งหนึ่งสำรวจอายุ (ปี) ของผู้ที่มาใช้บริการที่ร้านในหนึ่งวัน ได้ข้อมูลดังนี้

1	27	2	20	27	23	31	30	9
29	31	8	28	25	26	40	37	23
34	49	52	31	1	4	5	58	28
57	31	32	3	4	25	31	29	57
44	2	35	24	4	30	56	63	48

เรียงข้อมูลจากน้อยไปมากได้ดังนี้

1	1	2	2	3	4	4	4	5
8	9	20	23	23	24	25	25	26
27	27	28	28	29	29	30	30	31
31	31	31	31	32	34	35	37	40
44	48	49	52	56	57	57	58	63

จากข้อมูลข้างต้น สามารถเขียนแผนภาพลำต้นและใบได้ดังนี้

0	1 1 2 2 3 4 4 4 5 8 9
1	
2	0 3 3 4 5 5 6 7 7 8 8 9 9
3	0 0 1 1 1 1 1 2 4 5 7
4	0 4 8 9
5	2 6 7 7 8
6	3

จากแผนภาพลำต้นและใบ จะเห็นว่าในวันที่สำรวจข้อมูล ผู้ที่มาใช้บริการที่ร้านอาหารแห่งนี้มีอายุมากที่สุด 63 ปี และมีอายุน้อยที่สุด 1 ปี โดยผู้ที่มาใช้บริการที่ร้านอาหารแห่งนี้อยู่ในช่วงอายุ 20 – 29 ปี มากที่สุด

ข้อสังเกต ตัวเลขในส่วนลำต้นจะต้องเขียนเรียงลำดับให้ครบทุกตัว แต่ไม่ต้องเขียนตัวเลขในส่วนใบหากไม่มีข้อมูลนั้น เช่น จากสถานการณ์ข้างต้น จะเห็นว่าไม่มีข้อมูลในช่วงอายุ 10 – 19 ปี แต่ยังคงปรากฏ 1 ในส่วนลำต้น

นอกจากนี้สามารถเขียนแผนภาพลำต้นและใบในการนำเสนอข้อมูลสองชุดในแผนภาพเดียวกันได้ โดยใช้ส่วนลำต้นร่วมกัน เพื่อใช้เปรียบเทียบข้อมูลทั้งสองชุด

พิจารณาสถานการณ์ต่อไปนี้
คะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 ในโรงเรียนขนาดเล็กแห่งหนึ่ง ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงข้อมูลจากน้อยไปมาก แสดงได้ดังนี้

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1

56	60	65	69	73	73	75	78
83	85	87	87	89	93	94	97

นักเรียนชั้นมัธยมศึกษาปีที่ 6/2

57	63	64	65	68	68	70	72
73	73	75	78	81	83	87	98

จากข้อมูลข้างต้น สามารถเขียนแผนภาพลำต้นและใบได้ดังนี้

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1		นักเรียนชั้นมัธยมศึกษาปีที่ 6/2
6	5	7
9 5 0	6	3 4 5 8 8
8 5 3 3	7	0 2 3 3 5 8
9 7 7 5 3	8	1 3 7
7 4 3	9	8

จากแผนภาพลำต้นและใบ จะเห็นว่า

- คะแนนสอบสูงสุดวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 คือ 97 และ 98 คะแนน ตามลำดับ
- คะแนนสอบต่ำสุดวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 คือ 56 และ 57 คะแนน ตามลำดับ
- จำนวนนักเรียนที่ได้คะแนนสอบวิชาภาษาอังกฤษตั้งแต่ 80 คะแนนขึ้นไป ของชั้นมัธยมศึกษาปีที่ 6/1 มากกว่าชั้นมัธยมศึกษาปีที่ 6/2

แผนภาพกล่อง

แผนภาพกล่อง (box plot) เป็นการนำเสนอข้อมูลเชิงปริมาณที่แสดงตำแหน่งสำคัญของข้อมูล ซึ่งประกอบด้วยค่าต่ำสุด ค่าสูงสุด และควอร์ไทล์ (quartile) นอกจากนี้ แผนภาพกล่องสามารถใช้ในการตรวจสอบว่ามีข้อมูลที่แตกต่างไปจากข้อมูลส่วนใหญ่หรือไม่ โดยจะเรียกข้อมูลดังกล่าวว่า **ค่านอกเกณฑ์ (outlier)**

ขั้นตอนการเขียนแผนภาพกล่อง

1. เรียงข้อมูลจากน้อยไปมาก จากนั้นหาค่าต่ำสุดและค่าสูงสุดของข้อมูล
2. หาควอร์ไทล์ที่ 1 (Q_1) ควอร์ไทล์ที่ 2 (Q_2) และควอร์ไทล์ที่ 3 (Q_3) โดยที่
   - ควอร์ไทล์ที่ 1 (Q_1) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณหนึ่งในสี่ของจำนวนข้อมูลทั้งหมด
   - ควอร์ไทล์ที่ 2 (Q_2) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณครึ่งหนึ่งของจำนวนข้อมูลทั้งหมด หรือค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมด
   - ควอร์ไทล์ที่ 3 (Q_3) คือค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณสามในสี่ของจำนวนข้อมูลทั้งหมด

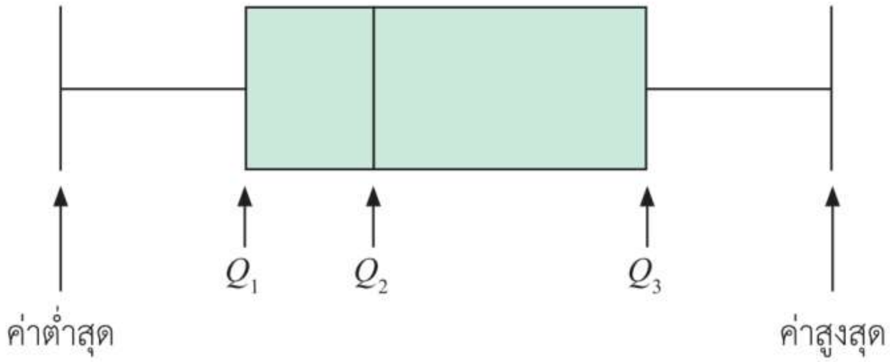
   ถ้า n เป็นจำนวนข้อมูลทั้งหมด สามารถหาตำแหน่งของควอร์ไทล์ได้ดังนี้

   Q_1 อยู่ในตำแหน่งที่ $\dfrac{n+1}{4}$

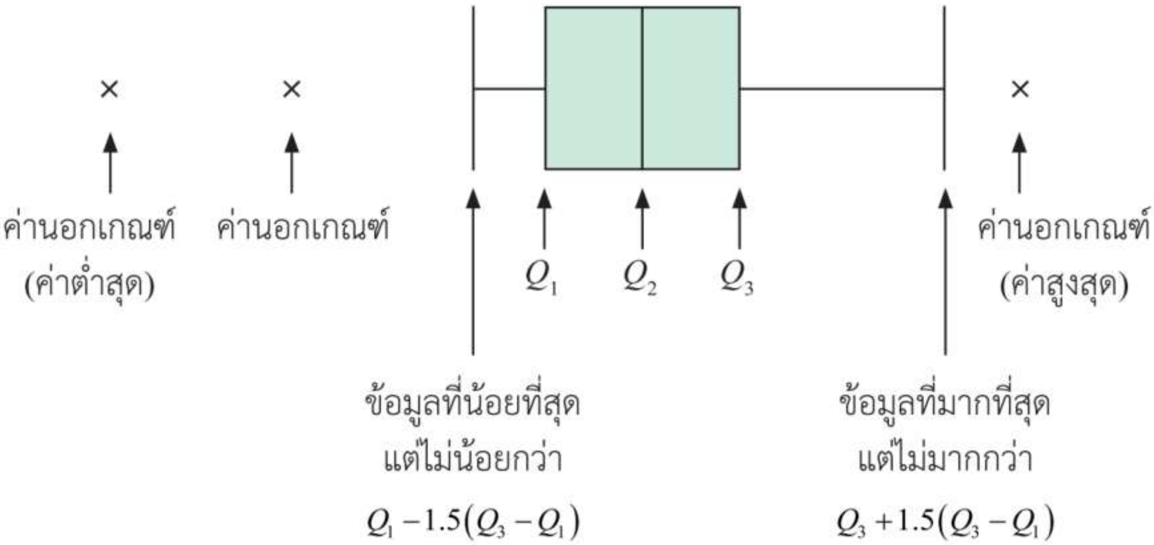
   Q_2 อยู่ในตำแหน่งที่ $\dfrac{2(n+1)}{4}$

   และ Q_3 อยู่ในตำแหน่งที่ $\dfrac{3(n+1)}{4}$

3. หาค่า $Q_1 - 1.5(Q_3 - Q_1)$ และ $Q_3 + 1.5(Q_3 - Q_1)$
4. พิจารณาว่าชุดข้อมูลมีค่านอกเกณฑ์หรือไม่ โดยในที่นี้ค่านอกเกณฑ์คือข้อมูลที่มีค่าน้อยกว่า $Q_1 - 1.5(Q_3 - Q_1)$ หรือข้อมูลที่มีค่ามากกว่า $Q_3 + 1.5(Q_3 - Q_1)$
5. ตัวอย่างการเขียนแผนภาพกล่อง
   - กรณีที่ไม่มีค่านอกเกณฑ์



   - กรณีที่มีค่านอกเกณฑ์



ค่านอกเกณฑ์อาจเป็นค่าจริงที่เกิดขึ้นตามธรรมชาติหรืออาจเกิดจากความคลาดเคลื่อนจากการวัดหรือเก็บข้อมูล ในทางปฏิบัติอาจไม่สามารถล่วงรู้ได้ว่าค่านอกเกณฑ์ที่ได้เกิดจากการวัดหรือเก็บข้อมูลที่ผิดพลาดหรือไม่

ตัวอย่างที่ 5

คะแนนสอบวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 27 คน ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงข้อมูลจากน้อยไปมาก แสดงได้ดังนี้

59	60	61	63	65	66	66	66	68
69	69	70	71	72	72	75	75	75
76	79	81	88	88	89	90	92	97

จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้

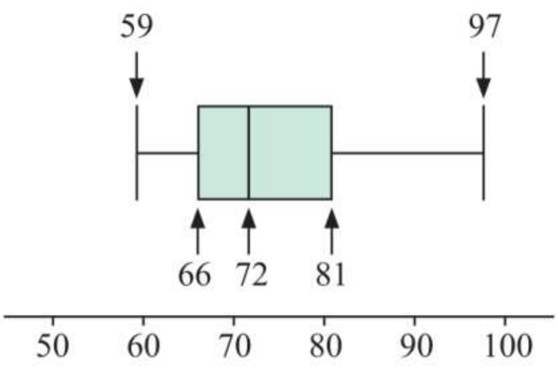
วิธีทำ จากโจทย์ มีข้อมูลทั้งหมด 27 ตัว เขียนแผนภาพกล่องได้ตามขั้นตอนต่อไปนี้

1. หาค่าต่ำสุดของข้อมูล ซึ่งคือ 59 และหาค่าสูงสุดของข้อมูล ซึ่งคือ 97
2. หา Q_1, Q_2 และ Q_3 ได้ดังนี้

   Q_1 อยู่ในตำแหน่งที่ $\frac{27+1}{4} = 7$ ดังนั้น $Q_1 = 66$

   Q_2 อยู่ในตำแหน่งที่ $\frac{2(27+1)}{4} = 14$ ดังนั้น $Q_2 = 72$

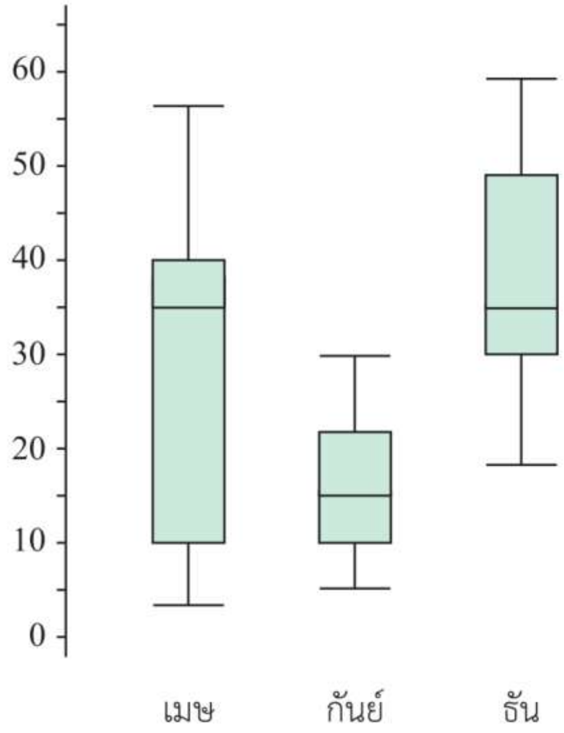
   Q_3 อยู่ในตำแหน่งที่ $\frac{3(27+1)}{4} = 21$ ดังนั้น $Q_3 = 81$
3. หาค่า $Q_1 - 1.5(Q_3 - Q_1)$ ซึ่งคือ $66 - 1.5(81-66) = 43.5$ และหาค่า $Q_3 + 1.5(Q_3 - Q_1)$ ซึ่งคือ $81 + 1.5(81-66) = 103.5$
4. เนื่องจากไม่มีข้อมูลที่มีค่าน้อยกว่า 43.5 หรือมากกว่า 103.5 ดังนั้น ข้อมูลชุดนี้ไม่มีค่านอกเกณฑ์
5. เขียนแผนภาพกล่องได้ดังนี้



แผนภาพกล่องอาจเขียนแสดงในแนวตั้งได้ ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 6

ระยะเวลา (นาที) ในการใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกของเมษ กันย์ และธัน ในแต่ละครั้งในเวลาหนึ่งสัปดาห์ แสดงด้วยแผนภาพกล่องได้ดังนี้



จงพิจารณาว่าข้อความต่อไปนี้เป็นจริง เป็นเท็จ หรือไม่สามารถสรุปได้ว่าเป็นจริงหรือเท็จ พร้อมทั้งให้เหตุผลประกอบ

1) ในหนึ่งสัปดาห์ธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในแต่ละครั้งเป็นระยะเวลานานกว่ากันย์
2) จำนวนครั้งที่กันย์ใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในหนึ่งสัปดาห์น้อยกว่าเมษและธัน
3) ควอร์ไทล์ที่ 2 ของระยะเวลาที่เมษและธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในแต่ละครั้งในหนึ่งสัปดาห์เท่ากัน

วิธีทำ 1) จากแผนภาพกล่อง จะได้ว่า
ระยะเวลาที่น้อยที่สุดที่ธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกประมาณ 18 นาที
ในขณะที่ระยะเวลาที่มากที่สุดที่กันย์ใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกคือ 30 นาที
แสดงว่ามีการโทรออกที่กันย์ใช้ระยะเวลานานกว่าธัน
ดังนั้น ข้อความ “ในหนึ่งสัปดาห์ธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในแต่ละครั้งเป็นระยะเวลานานกว่ากันย์” เป็นเท็จ

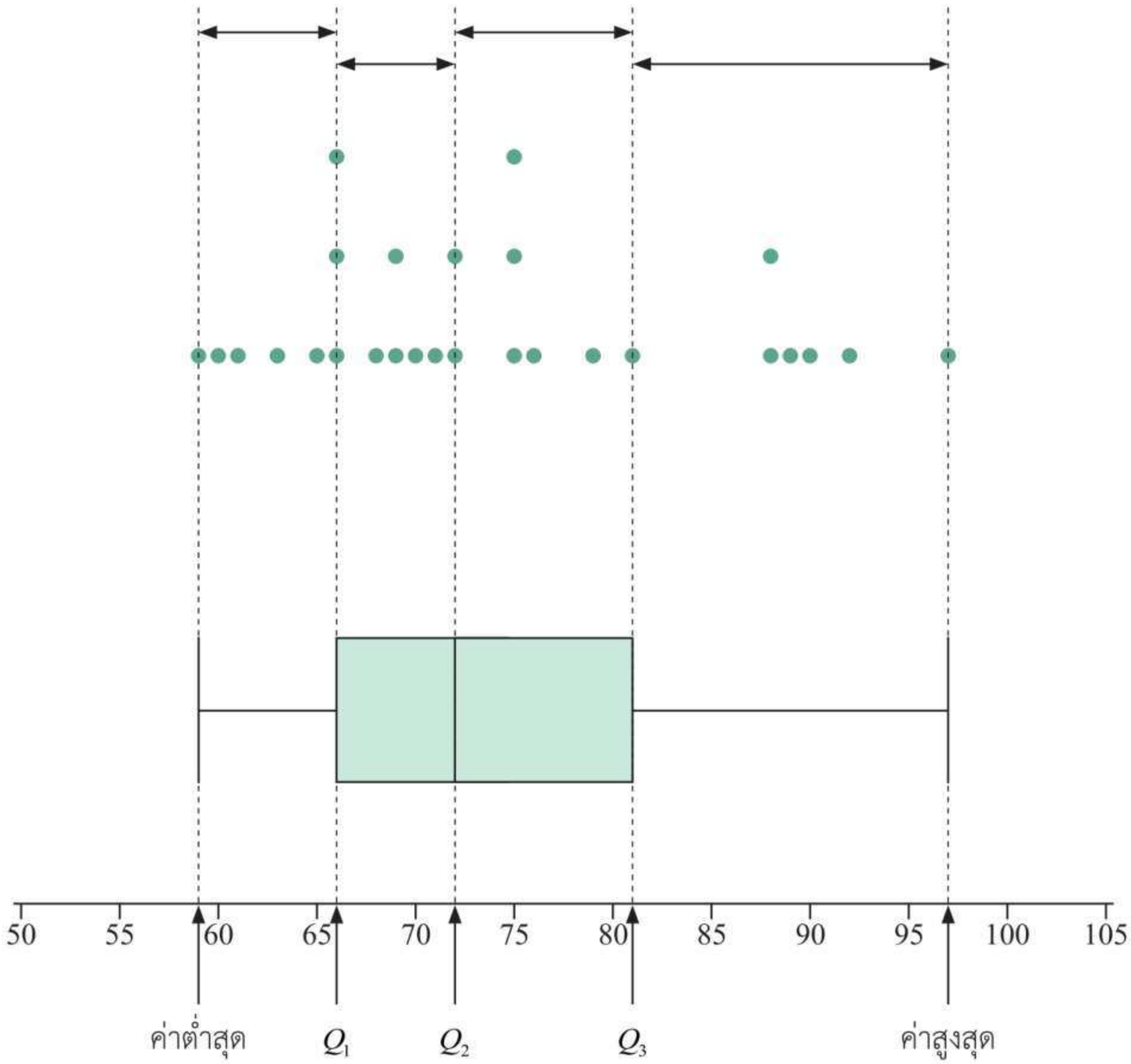
2) เนื่องจากแผนภาพกล่องแสดงระยะเวลาในการใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกของทั้งสามคน ซึ่งไม่สามารถพิจารณาจำนวนครั้งของการโทรออกของทั้งสามคนได้
ดังนั้น ไม่สามารถสรุปได้ว่าข้อความ “จำนวนครั้งที่กันย์ใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในหนึ่งสัปดาห์น้อยกว่าเมษและธัน” เป็นจริงหรือเท็จ

3) จากแผนภาพกล่อง จะได้ว่า
ควอร์ไทล์ที่ 2 ของระยะเวลาที่เมษและธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในแต่ละครั้งในหนึ่งสัปดาห์เท่ากัน คือ 35 นาที
ดังนั้น ข้อความ “ควอร์ไทล์ที่ 2 ของระยะเวลาที่เมษและธันใช้โทรศัพท์เคลื่อนที่สำหรับโทรออกในแต่ละครั้งในหนึ่งสัปดาห์เท่ากัน” เป็นจริง ■

แผนภาพกล่องกับการกระจายของข้อมูล

นอกจากแผนภาพกล่องจะสามารถใช้ในการตรวจสอบว่าชุดข้อมูลมีค่านอกเกณฑ์หรือไม่ ยังสามารถใช้ในการอธิบายลักษณะการกระจายของข้อมูลได้อีกด้วย การกระจายของข้อมูลจะทำให้เห็นว่าโดยภาพรวมแล้ว ข้อมูลมีการเกาะกลุ่มกันหรือไม่ ถ้าข้อมูลมีการกระจายมาก แสดงว่าข้อมูลมีค่าแตกต่างกันมากหรือข้อมูลไม่เกาะกลุ่มกัน แต่ถ้าข้อมูลมีการกระจายน้อย แสดงว่าข้อมูลมีค่าใกล้เคียงกันมากหรือข้อมูลเกาะกลุ่มกัน

จากตัวอย่างที่ 5 สามารถเขียนแผนภาพจุดเพื่อเปรียบเทียบกับแผนภาพกล่อง ได้ดังนี้



จากแผนภาพ จะเห็นว่าข้อมูลแบ่งออกเป็น 4 ช่วง แต่ละช่วงมีจำนวนข้อมูลประมาณ 25% ของจำนวนข้อมูลทั้งหมด เมื่อพิจารณาความกว้างของแต่ละช่วง จะพบว่าช่วงจาก Q_1 ถึง Q_2 มีความกว้างน้อยที่สุด ในขณะที่ช่วงจาก Q_3 ถึงค่าสูงสุด มีความกว้างมากที่สุด ทั้ง ๆ ที่ทั้งสองช่วงมีจำนวนข้อมูลเท่ากัน แสดงว่าข้อมูลที่อยู่ระหว่าง Q_1 และ Q_2 มีการกระจายน้อยที่สุด แต่ข้อมูลที่อยู่ระหว่าง Q_3 ถึงค่าสูงสุด มีการกระจายมากที่สุด

ตัวอย่างที่ 7

ข้อมูลจำนวนครั้งของการทำธุรกรรมผ่านเครือข่ายอินเทอร์เน็ตของครูในโรงเรียนแห่งหนึ่งในหนึ่งเดือน แสดงได้ดังนี้

0	0	0	0	2	2	2	3
3	3	3	3	3	4	5	5
5	5	5	6	6	6	6	6
7	9	10	11	12	12	14	

จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูล พร้อมทั้งอธิบายลักษณะการกระจายของข้อมูลชุดนี้

วิธีทำ จากโจทย์ มีข้อมูลทั้งหมด 31 ตัว เขียนแผนภาพกล่องได้ตามขั้นตอนต่อไปนี้

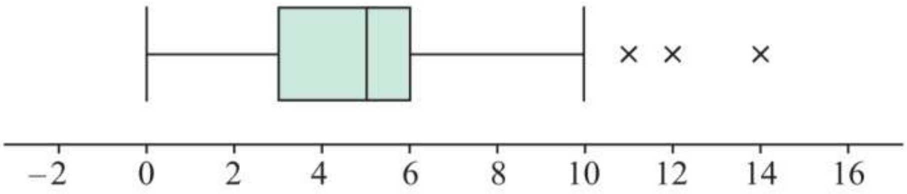
1. หาค่าต่ำสุดของข้อมูล ซึ่งคือ 0 และหาค่าสูงสุดของข้อมูล ซึ่งคือ 14
2. หา Q_1, Q_2 และ Q_3 ได้ดังนี้

   Q_1 อยู่ในตำแหน่งที่ $\frac{31+1}{4} = 8$ ดังนั้น $Q_1 = 3$

   Q_2 อยู่ในตำแหน่งที่ $\frac{2(31+1)}{4} = 16$ ดังนั้น $Q_2 = 5$

   Q_3 อยู่ในตำแหน่งที่ $\frac{3(31+1)}{4} = 24$ ดังนั้น $Q_3 = 6$
3. หาค่า $Q_1 - 1.5(Q_3 - Q_1)$ ซึ่งคือ $3 - 1.5(6-3) = -1.5$ และหาค่า $Q_3 + 1.5(Q_3 - Q_1)$ ซึ่งคือ $6 + 1.5(6-3) = 10.5$

4. จากข้อมูลข้างต้น มี 11, 12 และ 14 มากกว่า 10.5 แต่ไม่มีข้อมูลที่มีค่าน้อยกว่า -1.5 ดังนั้น ค่านอกเกณฑ์ ได้แก่ 11, 12 และ 14
5. เขียนแผนภาพกล่องได้ดังนี้



จากแผนภาพกล่อง จะเห็นว่าข้อมูลในช่วง 6 ถึง 10 มีการกระจายมากที่สุด รองลงมาคือข้อมูลในช่วง 0 ถึง 3 และข้อมูลในช่วง 5 ถึง 6 มีการกระจายน้อยที่สุด ■

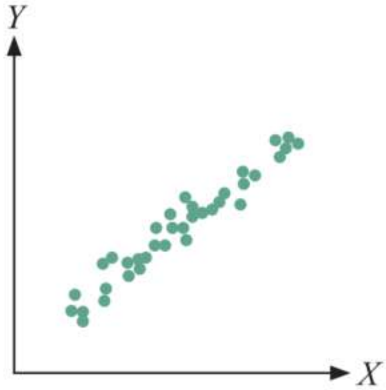
หมายเหตุ จากตัวอย่างข้างต้น เนื่องจากค่านอกเกณฑ์ไม่อยู่ในช่วง 6 ถึง 10 ดังนั้น ข้อมูลที่อยู่ในช่วง 6 ถึง 10 มีจำนวนข้อมูลไม่ถึง 25% ของจำนวนข้อมูลทั้งหมด

แผนภาพการกระจาย

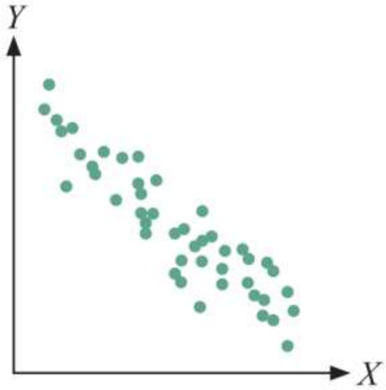
การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณข้างต้นพิจารณาเฉพาะข้อมูลเชิงปริมาณของตัวแปรเพียงหนึ่งตัว แต่ในหัวข้อนี้จะพิจารณาข้อมูลเชิงปริมาณของตัวแปรสองตัวว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นกันหรือไม่ เนื่องจากเมื่อมีตัวแปรสองตัว ไม่จำเป็นที่ตัวหนึ่งต้องเป็นตัวแปรต้นและอีกตัวต้องเป็นตัวแปรตาม การตรวจสอบความสัมพันธ์เชิงเส้นของตัวแปรทั้งสองจะพิจารณาโดยใช้แผนภาพการกระจาย

แผนภาพการกระจาย (scatter plot) คือแผนภาพที่เกิดจากการลงจุดที่แสดงค่าของตัวแปรคู่หนึ่ง รูปแบบการกระจายของจุดต่าง ๆ ที่ปรากฏในแผนภาพจะแสดงถึงรูปแบบความสัมพันธ์ระหว่างสองตัวแปรนั้น

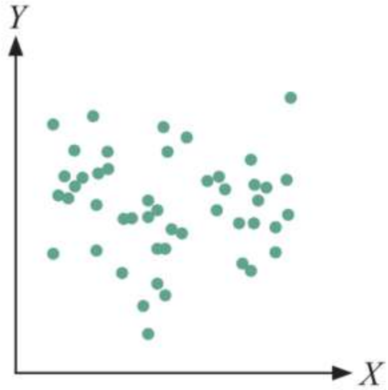
พิจารณาแผนภาพการกระจายต่อไปนี้



รูปที่ 1



รูปที่ 2



รูปที่ 3

เมื่อกำหนดให้

รูปที่ 1 ค่าบนแกน X และ Y แทนพื้นที่ห้องชุดและราคาห้องชุดของคอนโดมิเนียมในย่านธุรกิจแห่งหนึ่งในกรุงเทพมหานคร ตามลำดับ

รูปที่ 2 ค่าบนแกน X และ Y แทนอายุการใช้งานและมูลค่าของรถยนต์ยี่ห้อหนึ่ง ตามลำดับ

รูปที่ 3 ค่าบนแกน X และ Y แทนอายุและรายได้ของประชากรวัยทำงานในหมู่บ้านแห่งหนึ่ง ตามลำดับ

จากรูปที่ 1 จะเห็นว่าเมื่อพื้นที่ห้องชุดมากขึ้น ราคาห้องชุดจะมากขึ้นด้วย ในกรณีนี้จะกล่าวว่าตัวแปรทั้งสองมีความสัมพันธ์ในทิศทางเดียวกัน

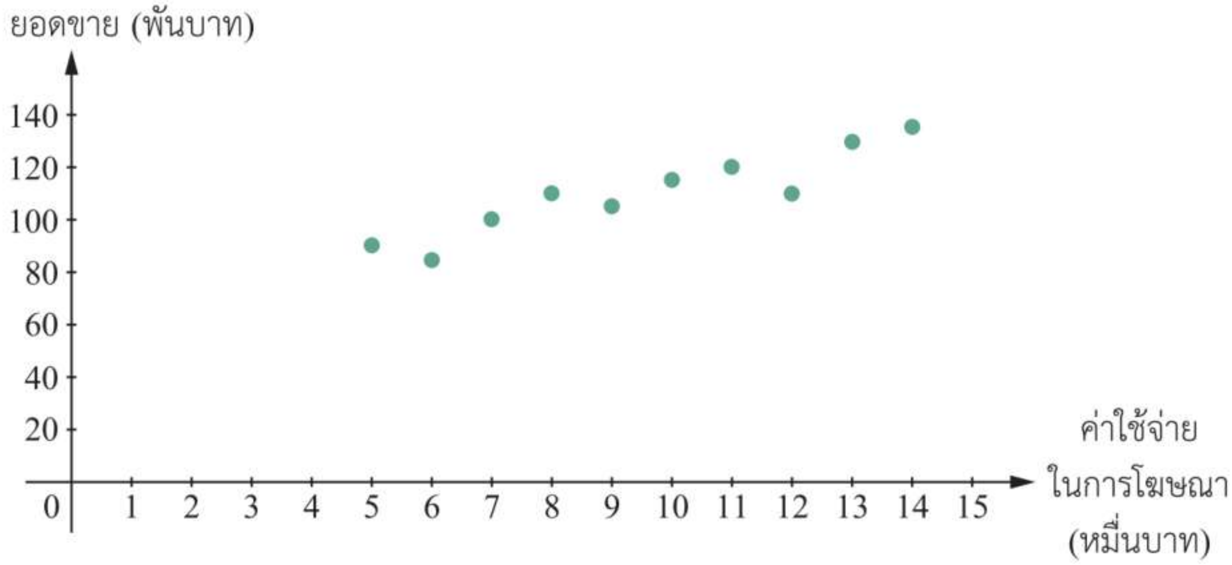
จากรูปที่ 2 จะเห็นว่าเมื่ออายุการใช้งานมากขึ้น มูลค่าของรถยนต์จะยิ่งน้อยลง ในกรณีนี้จะกล่าวว่าตัวแปรทั้งสองมีความสัมพันธ์ในทิศทางตรงกันข้าม

จากรูปที่ 3 จะเห็นว่าเมื่ออายุมากขึ้น รายได้ของประชากรวัยทำงานในหมู่บ้านแห่งนี้ไม่ได้มากขึ้นหรือน้อยลงตาม ในกรณีนี้จะกล่าวว่าตัวแปรทั้งสองไม่มีความสัมพันธ์เชิงเส้น

หมายเหตุ ในที่นี้จะพิจารณาเฉพาะความสัมพันธ์เชิงเส้นเท่านั้น

ตัวอย่างที่ 8

บริษัทแห่งหนึ่งสำรวจยอดขายของผลิตภัณฑ์ที่มีค่าใช้จ่ายในการโฆษณาผลิตภัณฑ์นั้นแตกต่างกันในพื้นที่ต่าง ๆ ทั่วประเทศไทย จำนวน 10 แห่ง ได้ผลสำรวจดังนี้



จงพิจารณาว่าค่าใช้จ่ายในการโฆษณาและยอดขายของผลิตภัณฑ์มีความสัมพันธ์กันหรือไม่ อย่างไร

วิธีทำ จากแผนภาพการกระจาย จะเห็นว่าเมื่อค่าใช้จ่ายในการโฆษณามากขึ้น ยอดขายของผลิตภัณฑ์จะมีแนวโน้มมากขึ้นด้วย ดังนั้น จึงกล่าวได้ว่าค่าใช้จ่ายในการโฆษณาและยอดขายของผลิตภัณฑ์มีความสัมพันธ์ในทิศทางเดียวกัน ■

แบบฝึกหัด 3.2

1. คะแนนสอบย่อยวิชาคณิตศาสตร์ซึ่งมีคะแนนเต็ม 10 คะแนน ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 30 คน แสดงได้ดังนี้

คะแนน	0	1	2	3	4	5	6	7	8	9	10
ความถี่	1	0	2	3	2	5	3	2	6	5	1

จงเขียนฮิสโทแกรมเพื่อนำเสนอข้อมูลชุดนี้

2. โรงเรียนประจำจังหวัดแห่งหนึ่งในภาคใต้ได้จัดโครงการตรวจสุขภาพครู ประจำปี 2562 โดยผลการตรวจระดับน้ำตาลในเลือด (มิลลิกรัมต่อเดซิลิตร) ของครูทั้งหมด 80 คน แสดงได้ดังนี้

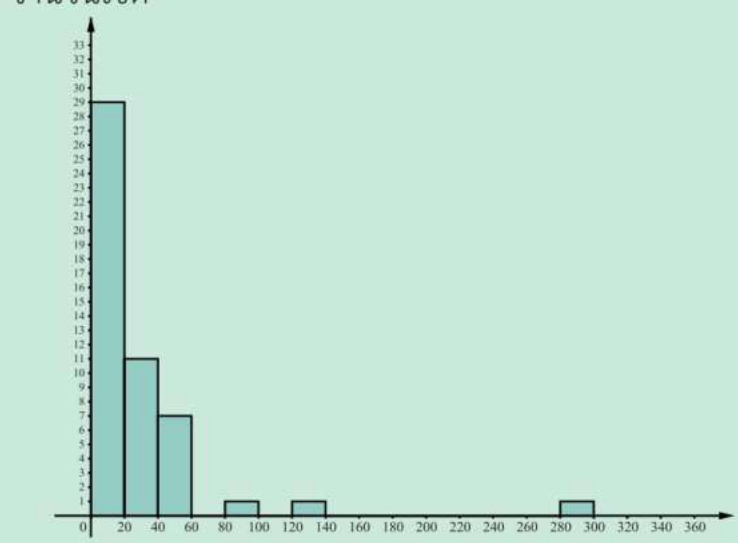
ระดับน้ำตาลในเลือด (มิลลิกรัมต่อเดซิลิตร)	ความถี่สะสม
61 – 67	3
68 – 74	6
75 – 81	9
82 – 88	24
89 – 95	47
96 – 102	63
103 – 109	76
110 – 116	80

1) ครูมีระดับน้ำตาลในเลือดอยู่ในช่วงใดมากที่สุด
2) จงเขียนฮิสโทแกรมจากข้อมูลที่กำหนดให้
3) ถ้าระดับน้ำตาลในเลือดอยู่ในช่วง 100 – 125 มิลลิกรัมต่อเดซิลิตร จะถือว่ามีความเสี่ยงต่อการเป็นโรคเบาหวาน จงพิจารณาว่าข้อความ “มีครูมากกว่าร้อยละ 42 ของครูทั้งหมดมีความเสี่ยงต่อการเป็นโรคเบาหวาน” เป็นจริงหรือไม่ เพราะเหตุใด

เสริมสมอง : การตรวจวัดระดับน้ำตาลในเลือด

การตรวจวัดระดับน้ำตาลในเลือด (fasting blood sugar: FBS หรือ fasting blood glucose: FBG) คือการวัดปริมาณกลูโคสในเลือด โดยให้งดรับประทานอาหารและเครื่องดื่มอย่างน้อย 8 ชั่วโมง ก่อนการเจาะเลือด เพื่อคัดกรองและวินิจฉัยผู้ที่มีความเสี่ยงเป็นโรคเบาหวาน นอกจากนี้ยังช่วยติดตามและประเมินผลการรักษาเพื่อป้องกันไม่ให้ระดับน้ำตาลในเลือดสูงเกินไปในผู้ป่วยโรคเบาหวาน โดยคนปกติจะมีระดับน้ำตาลในเลือดประมาณ 70 – 100 มิลลิกรัมต่อเดซิลิตร แต่ถ้าระดับน้ำตาลในเลือดมากกว่า 126 มิลลิกรัมต่อเดซิลิตร อาจบ่งชี้ว่าเป็นโรคเบาหวาน

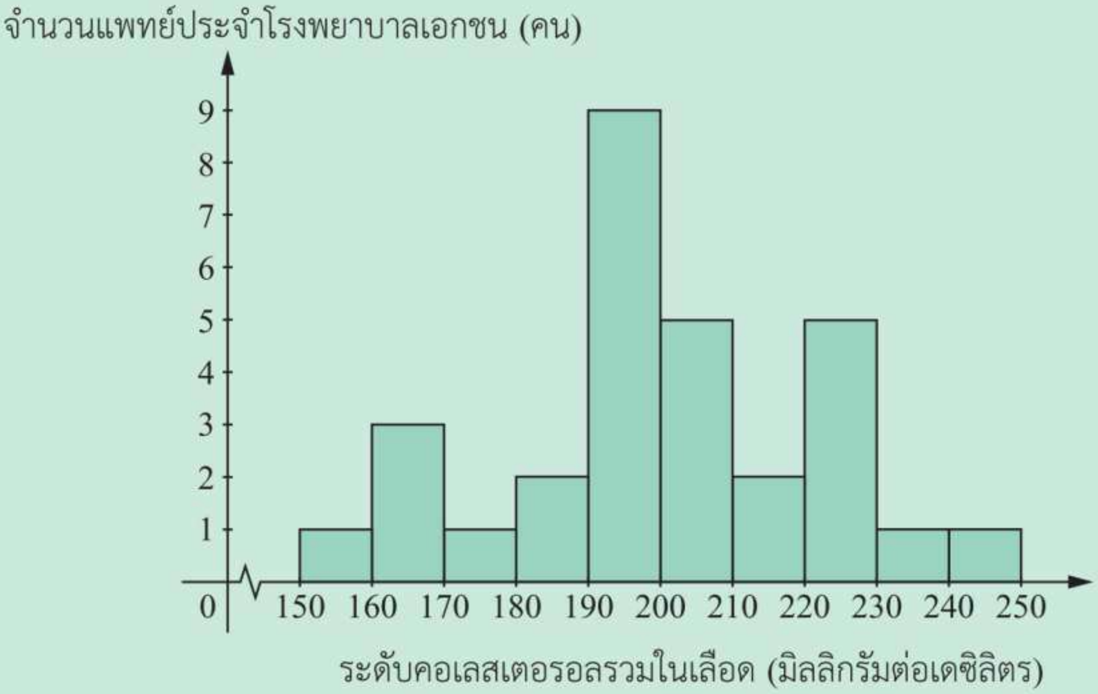
3. อัตราส่วนพื้นที่สีเขียว 10 ประเภท ต่อจำนวนประชากร (ตารางเมตรต่อคน) ของแต่ละเขตในกรุงเทพมหานคร ณ วันที่ 22 กันยายน พ.ศ. 2562 แสดงด้วยฮิสโทแกรมได้ดังนี้

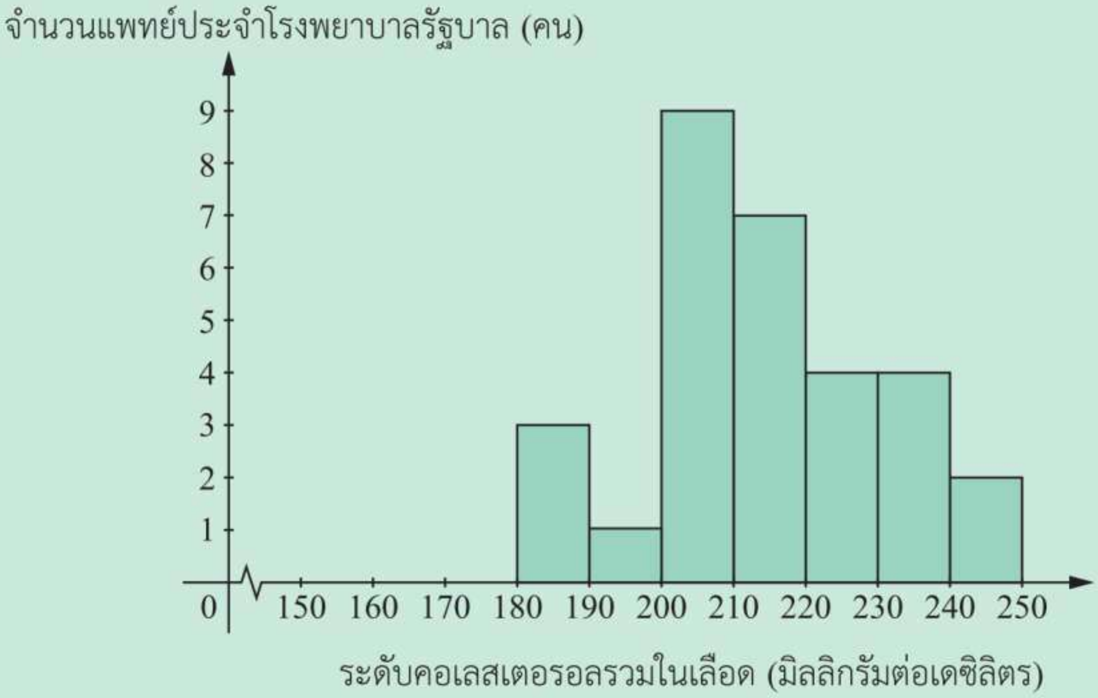


อัตราส่วนของพื้นที่สีเขียว 10 ประเภท ต่อจำนวนประชากร (ตารางเมตรต่อคน)

1) มีกี่เขตที่อัตราส่วนพื้นที่สีเขียว 10 ประเภท ต่อจำนวนประชากรน้อยกว่า 40 ตารางเมตรต่อคน
2) มีกี่เขตที่อัตราส่วนพื้นที่สีเขียว 10 ประเภท ต่อจำนวนประชากรไม่น้อยกว่า 80 ตารางเมตรต่อคน
3) ถ้าเขตที่มีอัตราส่วนพื้นที่สีเขียว 10 ประเภท ต่อจำนวนประชากรมากที่สุดมีประชากร 182,235 คน จะสามารถสรุปได้หรือไม่ว่ามีพื้นที่สีเขียวในเขตนั้นไม่น้อยกว่า 50,000,000 ตารางเมตร

4. ระดับคอเลสเตอรอลรวม (total cholesterol) ในเลือด (มิลลิกรัมต่อเดซิลิตร) ของแพทย์ประจำโรงพยาบาลเอกชนแห่งหนึ่งและแพทย์ประจำโรงพยาบาลรัฐบาลแห่งหนึ่งในกรุงเทพมหานคร โดยสุ่มตัวอย่างมาโรงพยาบาลละ 30 คน แสดงด้วยฮิสโทแกรมได้ดังนี้





1) แพทย์ประจำโรงพยาบาลเอกชนที่สุ่มมามีระดับคอเลสเตอรอลรวมในเลือดอยู่ในช่วงใดมากที่สุด และมีแพทย์อยู่ในช่วงดังกล่าวกี่คน
2) ระดับคอเลสเตอรอลรวมในเลือดของแพทย์ประจำโรงพยาบาลรัฐบาลที่สุ่มมามีค่าสูงสุดอยู่ในช่วงใด และมีแพทย์อยู่ในช่วงดังกล่าวกี่คน
3) ถ้าค่าปกติของระดับคอเลสเตอรอลรวมในเลือดควรน้อยกว่า 200 มิลลิกรัมต่อเดซิลิตร
   3.1) แพทย์ประจำโรงพยาบาลเอกชนที่มีระดับคอเลสเตอรอลรวมในเลือดสูงกว่าค่าปกติคิดเป็นร้อยละเท่าใดของจำนวนแพทย์ประจำโรงพยาบาลเอกชนที่สุ่มมาทั้งหมด
   3.2) แพทย์ประจำโรงพยาบาลเอกชนและแพทย์ประจำโรงพยาบาลรัฐบาลที่มีระดับคอเลสเตอรอลรวมในเลือดสูงกว่าค่าปกติคิดเป็นร้อยละเท่าใดของจำนวนแพทย์ที่สุ่มมาทั้งหมด
   3.3) จากกลุ่มแพทย์ที่สุ่มมา แพทย์ประจำโรงพยาบาลเอกชนที่มีระดับคอเลสเตอรอลรวมในเลือดสูงกว่าค่าปกติมีจำนวนมากหรือน้อยกว่าแพทย์ประจำโรงพยาบาลรัฐบาลที่มีระดับคอเลสเตอรอลรวมในเลือดสูงกว่าค่าปกติ และมีจำนวนต่างกันเท่าใด
4) จงสรุปลักษณะที่น่าสนใจเกี่ยวกับระดับคอเลสเตอรอลรวมในเลือดของแพทย์ประจำโรงพยาบาลเอกชนและแพทย์ประจำโรงพยาบาลรัฐบาลที่สุ่มมา

เสริมสมอง : คอเลสเตอรอล

คอเลสเตอรอลเป็นลิพิดชนิดหนึ่งซึ่งมีความสำคัญต่อร่างกาย เป็นส่วนประกอบของเยื่อหุ้มเซลล์ และเป็นสารตั้งต้นสำหรับการสังเคราะห์ฮอร์โมนบางชนิด เช่น เทสโทสเทอโรน อีสโทรเจน คอเลสเตอรอลในระดับที่เหมาะสมจะทำให้ระบบต่าง ๆ ในร่างกายทำงานได้ปกติ แต่หากร่างกายมีระดับคอเลสเตอรอลในเลือดที่สูงเกินไปจะทำให้เกิดการสะสมของคอเลสเตอรอลที่ผนังหลอดเลือด ทำให้หลอดเลือดแข็งตัวหรือตีบ อาจก่อให้เกิดภาวะหัวใจขาดเลือด หรือหลอดเลือดในสมองตีบ อาจทำให้เสียชีวิตได้ คนปกติควรมีระดับคอเลสเตอรอลในเลือดน้อยกว่า 200 มิลลิกรัมต่อเดซิลิตร

5. จำนวนภาพยนตร์ (เรื่อง) ที่นักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 50 คน ชมในหนึ่งปี เป็นดังนี้

33	17	15	18	20	10	15	5	22	25
22	6	16	24	25	17	7	20	19	3
12	16	21	32	16	16	23	25	28	25
26	27	9	17	5	20	15	16	7	19
32	17	16	7	18	26	28	18	16	10

1) จงเขียนแผนภาพจุดและแผนภาพลำต้นและใบจากข้อมูลที่กำหนดให้
2) นักเรียนที่ชมภาพยนตร์มากกว่า 12 เรื่องในหนึ่งปี คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด

6. คะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนในโรงเรียนแห่งหนึ่ง จำนวน 135 คน ซึ่งมีคะแนนเต็มวิชาละ 50 คะแนน แสดงด้วยตารางความถี่ได้ดังนี้

คะแนน	จำนวนนักเรียน (คน)	
	วิชาคณิตศาสตร์	วิชาภาษาอังกฤษ
50	5	2
49	12	3
48	18	5
47	19	0
46	24	12
45	19	24
44	13	35
43	9	22
42	6	11
41	4	9
40	3	6
39	1	4
38	2	2

ถ้าโรงเรียนต้องการมอบรางวัลให้นักเรียนที่ได้คะแนนสอบวิชาคณิตศาสตร์หรือวิชาภาษาอังกฤษตั้งแต่ร้อยละ 90 ขึ้นไปของคะแนนเต็มแต่ละวิชา โดยนักเรียน 1 คน สามารถรับได้มากกว่า 1 รางวัล จงหาว่ามีนักเรียนกี่คนที่ได้รางวัลในแต่ละวิชา

7. คะแนนสอบวิชาภาษาไทยและวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 40 คน ซึ่งมีคะแนนเต็มวิชาละ 100 คะแนน แสดงได้ดังนี้

คะแนนสอบวิชาภาษาไทย

63	88	79	92	86	87	83	83	78	40
68	76	46	81	92	77	77	84	76	70
77	75	98	81	82	81	81	87	78	70
94	79	52	82	77	81	81	77	70	74

คะแนนสอบวิชาคณิตศาสตร์

69	78	79	90	56	67	73	59	65	45
47	55	59	82	95	85	75	74	65	81
57	66	63	71	49	52	70	60	50	70
64	78	52	82	77	81	80	77	70	74

1) จงเขียนแผนภาพลำต้นและใบเพื่อนำเสนอข้อมูลสองชุดนี้ในแผนภาพเดียวกัน
2) ถ้ากำหนดเกณฑ์ในการแบ่งระดับคะแนนของทั้งสองวิชา ดังนี้

เกรด	คะแนน
4	90 – 100
3	80 – 89
2	70 – 79
1	60 – 69
0	0 – 59

จงหาจำนวนนักเรียนที่ได้เกรด 4, 3, 2, 1 และ 0 ของแต่ละวิชา

3) จงหาว่ามีนักเรียนกี่คนที่คะแนนสอบวิชาคณิตศาสตร์ขาดไปเพียง 1 คะแนน ก็จะได้เกรดดีขึ้น
4) จากเกรดวิชาภาษาไทยและวิชาคณิตศาสตร์ของนักเรียนห้องนี้ จงพิจารณาว่านักเรียนห้องนี้ถนัดวิชาใดมากกว่ากัน

8. นภาพักตร์ได้จดบันทึกจำนวนรถจักรยานยนต์ที่มาจอดบริเวณหน้าบ้านในช่วงเวลา 08:00 – 09:00 น. ของแต่ละวัน เป็นเวลา 1 เดือน ได้ข้อมูลดังนี้

10	6	7	12	13	15	8	6	10	16
17	20	18	5	9	9	7	10	11	18
19	15	16	17	20	16	12	14	18	14
17									

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) ข้อมูลชุดนี้มีค่านอกเกณฑ์หรือไม่ ถ้ามีคือค่าใด
3) จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้
4) จากแผนภาพกล่องที่ได้ในข้อ 3) จงอธิบายลักษณะการกระจายของข้อมูลชุดนี้

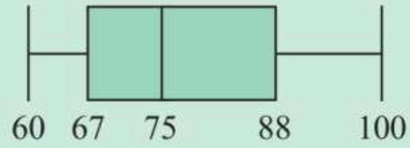
9. พิชพิชัยได้จดบันทึกความยาวของกล้วยทอด (เซนติเมตร) ที่สุ่มมาจากร้านค้าแห่งหนึ่งจำนวน 3 ถุง แต่ละถุงมีกล้วยทอด 9 ชิ้น ได้ข้อมูลดังนี้

5.0	7.0	6.0	6.5	3.6	5.4	5.6	8.0	7.5	6.0
5.8	6.9	7.3	7.5	7.7	7.0	6.8	6.3	5.9	6.1
7.2	6.4	5.5	8.0	5.8	6.7	7.2			

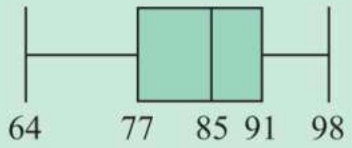
1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) ข้อมูลชุดนี้มีค่านอกเกณฑ์หรือไม่ ถ้ามีคือค่าใด
3) จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้
4) จากแผนภาพกล่องที่ได้ในข้อ 3) จงอธิบายลักษณะการกระจายของข้อมูลชุดนี้

10. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 แสดงด้วยแผนภาพกล่องได้ดังนี้

นักเรียนชั้นมัธยมศึกษาปีที่ 6/1



นักเรียนชั้นมัธยมศึกษาปีที่ 6/2



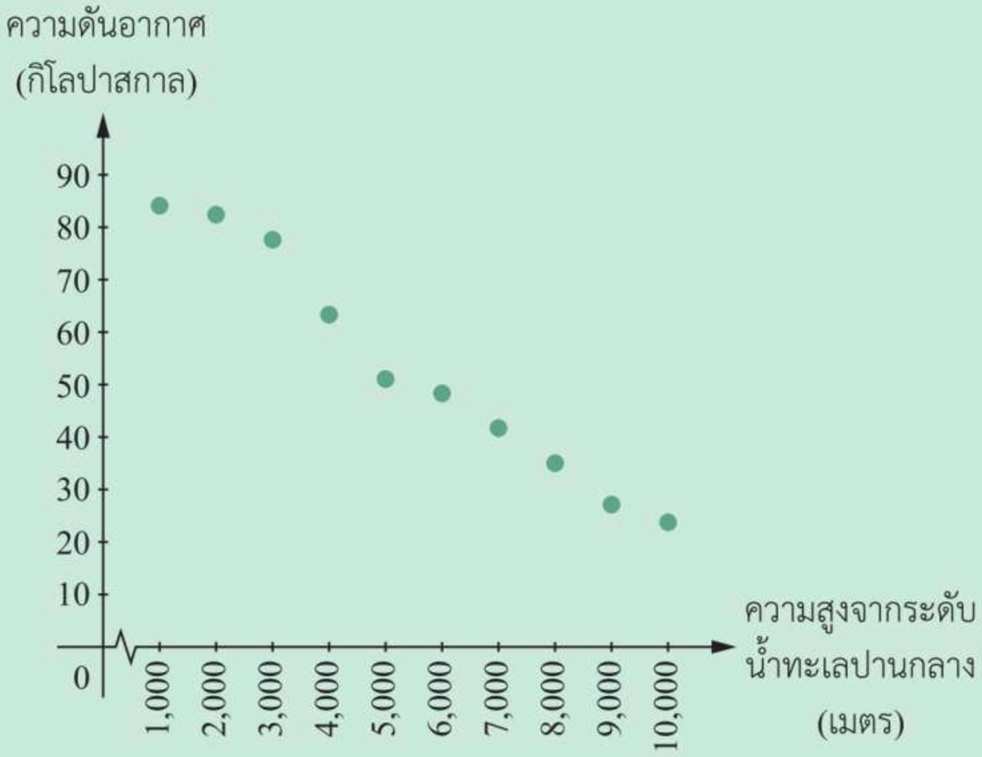
1) กลุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 ที่ได้คะแนนต่ำสุดมีจำนวนประมาณ 25% ของนักเรียนทั้งห้อง จงหาคะแนนต่ำสุดและคะแนนสูงสุดของนักเรียนกลุ่มนี้
2) มีนักเรียนชั้นมัธยมศึกษาปีที่ 6/2 กี่เปอร์เซ็นต์ที่ได้คะแนนมากกว่า 91 คะแนน
3) มีนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 กี่เปอร์เซ็นต์ที่ได้คะแนนน้อยกว่า 75 คะแนน
4) มีนักเรียนชั้นมัธยมศึกษาปีที่ 6/2 กี่เปอร์เซ็นต์ที่ได้คะแนนมากกว่า 77 คะแนน
5) กำหนดให้นักเรียนชั้นมัธยมศึกษาปีที่ 6/1 และ 6/2 มีจำนวนเท่ากัน และนักเรียนได้เกรด 4 ก็ต่อเมื่อนักเรียนได้คะแนนตั้งแต่ 80 คะแนนขึ้นไป จงพิจารณาว่าห้องใดน่าจะมีนักเรียนได้เกรด 4 มากกว่ากัน พร้อมทั้งให้เหตุผลประกอบ

11. ข้อมูลความสูง (เซนติเมตร) และน้ำหนัก (กิโลกรัม) ของนักเรียนที่สุ่มมาจำนวน 15 คน แสดงได้ดังนี้

ความสูง (เซนติเมตร)	168	152	155	149	145	154	146	160	162	152	163	157	164	141	145
น้ำหนัก (กิโลกรัม)	60	40	45	41	34	33	40	42	55	39	60	46	50	30	39

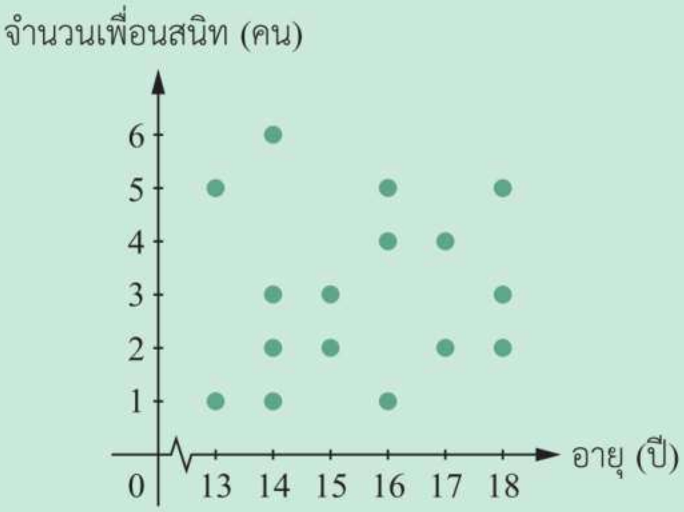
จงเขียนแผนภาพการกระจายของข้อมูลชุดนี้ พร้อมทั้งพิจารณาว่าความสูงและน้ำหนักของนักเรียนมีความสัมพันธ์กันหรือไม่ อย่างไร

12. ความดันอากาศ ณ ความสูงต่าง ๆ จากระดับน้ำทะเลปานกลาง แสดงด้วยแผนภาพการกระจายได้ดังนี้



จงพิจารณาว่าความดันอากาศและความสูงจากระดับน้ำทะเลปานกลางมีความสัมพันธ์กันหรือไม่ อย่างไร

13. จำนวนเพื่อนสนิทและอายุของนักเรียนที่สุ่มมาจำนวน 16 คน แสดงด้วยแผนภาพการกระจายได้ดังนี้



จงพิจารณาว่าจำนวนเพื่อนสนิทและอายุของนักเรียนมีความสัมพันธ์กันหรือไม่ อย่างไร

3.3 ค่าวัดทางสถิติ

ในการวิเคราะห์ข้อมูลโดยใช้วิธีการของสถิติศาสตร์เชิงพรรณนา นอกจากจะทำได้โดยการเขียนตารางความถี่หรือแผนภาพต่าง ๆ ตามที่ได้กล่าวมาแล้ว ยังสามารถใช้ค่าวัดทางสถิติซึ่งเป็นค่าที่ได้จากการนำข้อมูลทั้งหมดหรือข้อมูลบางส่วนมาคำนวณเพื่อใช้ในการวิเคราะห์ข้อมูล โดยค่าวัดทางสถิติจะช่วยให้เห็นภาพรวมของข้อมูลและช่วยให้สามารถจดจำข้อสรุปเกี่ยวกับข้อมูลนั้น ๆ ได้ง่ายขึ้น ซึ่งจะเป็นประโยชน์ในการนำไปประกอบการตัดสินใจหรือการวางแผนต่าง ๆ

ค่าวัดทางสถิติประกอบด้วยค่าสถิติและพารามิเตอร์ ซึ่งเป็นค่าที่วัดลักษณะโดยประมวลหรือคำนวณจากข้อมูลเหมือนกัน แต่แตกต่างกันตรงที่ค่าสถิติได้จากการพิจารณาข้อมูลของตัวอย่าง ในขณะที่พารามิเตอร์ได้จากการพิจารณาข้อมูลทั้งหมดของประชากร โดยค่าสถิติและพารามิเตอร์ที่จะศึกษาในหัวข้อนี้ ได้แก่ ค่ากลางของข้อมูล ค่าวัดการกระจาย และค่าวัดตำแหน่งที่ของข้อมูล

3.3.1 ค่ากลางของข้อมูล

ค่ากลางของข้อมูลมีหลายชนิด เช่น ค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม ซึ่งนักเรียนได้ศึกษามาแล้วในระดับมัธยมศึกษาตอนต้น ค่ากลางแต่ละชนิดต่างก็มีข้อดี ข้อเสีย และมีความเหมาะสมในการนำไปใช้ไม่เหมือนกัน ขึ้นอยู่กับลักษณะการแจกแจงของข้อมูลและวัตถุประสงค์ของผู้ใช้ข้อมูลนั้น ๆ

ในทางสถิติจะใช้ค่ากลางของข้อมูลเป็นตัวแทนของข้อมูลทั้งหมด เพื่อให้เข้าใจภาพรวมและสะดวกในการจดจำข้อสรุปเกี่ยวกับข้อมูลนั้น ๆ เช่น ผู้อำนวยการโรงเรียนแห่งหนึ่งต้องการทราบผลการเรียนวิชาคณิตศาสตร์ของนักเรียนในระดับชั้นต่าง ๆ ในปีที่ผ่านมา ผู้อำนวยการอาจไม่จำเป็นต้องทราบผลการเรียนวิชาคณิตศาสตร์ของนักเรียนแต่ละคนในแต่ละระดับชั้น แต่อาจพิจารณาเบื้องต้นจากค่าเฉลี่ยเลขคณิตของคะแนนวิชาคณิตศาสตร์ของนักเรียนแต่ละระดับชั้น

ค่าเฉลี่ยเลขคณิต

ค่าเฉลี่ยเลขคณิต (arithmetic mean) เป็นค่าที่หาได้จากการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลที่มี

ให้ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร
ค่าเฉลี่ยเลขคณิตของประชากร (population mean) เขียนแทนด้วย μ (อ่านว่า มิว) หาได้จาก

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N}$$

ให้ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง
ค่าเฉลี่ยเลขคณิตของตัวอย่าง (sample mean) เขียนแทนด้วย $\overline{x}$ (อ่านว่า เอ็กซ์บาร์) หาได้จาก

$$\overline{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

เพื่อความสะดวกจะใช้ตัวอักษรกรีกตัวพิมพ์ใหญ่ $\sum$ (อ่านว่า ซิกมา) เป็นสัญลักษณ์แสดงการบวก กล่าวคือ จะเขียนแทน $x_1 + x_2 + x_3 + \cdots + x_n$ ด้วยสัญลักษณ์ $\sum_{i=1}^{n} x_i$ (อ่านว่า ซัมเมชัน x_i เมื่อ i เท่ากับ 1 ถึง n) ดังนั้น สามารถเขียนสูตรของค่าเฉลี่ยเลขคณิตได้ดังนี้

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \quad \text{และ} \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

หมายเหตุ การหาค่าเฉลี่ยเลขคณิตจะใช้ได้กับข้อมูลเชิงปริมาณเท่านั้น เช่น สามารถหาค่าเฉลี่ยเลขคณิตของอายุ รายได้ ความสูง แต่จะไม่สามารถหาค่าเฉลี่ยเลขคณิตของข้อมูลเชิงคุณภาพ เช่น ยี่ห้อรถยนต์ เชื้อชาติ เบอร์โทรศัพท์

ตัวอย่างที่ 9

ทีมฟุตบอลชายของโรงเรียนแห่งหนึ่งมีสมาชิกทั้งหมด 24 คน โดยความสูง (เซนติเมตร) ของสมาชิกแต่ละคน แสดงได้ดังนี้

165	178	170	168	167	167	180	175
181	164	179	158	177	163	165	172
180	191	185	176	175	183	177	179

จงหาค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

วิธีทำ ให้ x_i แทนความสูงของนักฟุตบอลคนที่ i เมื่อ $i \in \{1, 2, 3, ..., 24\}$ เนื่องจากข้อมูลชุดนี้เป็นข้อมูลของประชากร จึงให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

จะได้
$$\mu = \frac{\sum_{i=1}^{24} x_i}{24} = \frac{4,175}{24} \approx 173.96$$

ดังนั้น ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้มีค่าประมาณ 173.96 เซนติเมตร ■

ข้อสังเกต ค่าเฉลี่ยเลขคณิตที่หาได้อาจไม่ใช่ค่าใดค่าหนึ่งของข้อมูลชุดนั้น เช่น จากตัวอย่างที่ 9 ค่าเฉลี่ยเลขคณิตของความสูงของสมาชิกทีมฟุตบอลของโรงเรียนมีค่าประมาณ 173.96 เซนติเมตร แต่ไม่มีสมาชิกคนใดมีความสูง 173.96 เซนติเมตร

ตัวอย่างที่ 10

โรงเรียนแห่งหนึ่งกำหนดว่านักเรียนจะได้เกรด 4 วิชาคณิตศาสตร์ ก็ต่อเมื่อนักเรียนได้คะแนนเฉลี่ยจากการสอบย่อย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน ถ้าค่าเฉลี่ยเลขคณิตจากการสอบย่อย 5 ครั้ง ของนักเรียนคนหนึ่งเท่ากับ 77 คะแนน จงหาว่าในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้จะต้องได้คะแนนอย่างน้อยเท่าใด จึงจะได้เกรด 4

วิธีทำ ให้ x_i แทนคะแนนสอบย่อยครั้งที่ i ของนักเรียนคนนี้ เมื่อ $i \in \{1, 2, 3, 4, 5, 6\}$

จากโจทย์ จะได้ $$\frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = 77$$

ดังนั้น $$x_1 + x_2 + x_3 + x_4 + x_5 = 385$$

ถ้าต้องการได้เกรด 4 วิชาคณิตศาสตร์ จะต้องได้คะแนนเฉลี่ยจากการสอบย่อย 6 ครั้ง ไม่ต่ำกว่า 80 คะแนน

นั่นคือ $$\frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \geq 80$$

$$\frac{385 + x_6}{6} \geq 80$$

$$x_6 \geq 95$$

ดังนั้น ในการสอบย่อยครั้งที่ 6 นักเรียนคนนี้จะต้องได้คะแนนอย่างน้อย 95 คะแนน จึงจะได้เกรด 4 วิชาคณิตศาสตร์ ■

ตัวอย่างที่ 11

คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่สุ่มตัวอย่างมาจากห้องหนึ่ง จำนวน 11 คน เป็นดังนี้

70 72 68 3 71 74 70 67 73 5 78

จงหา

1) ค่านอกเกณฑ์ของข้อมูลชุดนี้
2) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้
3) ค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มมา โดยไม่รวมค่านอกเกณฑ์

วิธีทำ 1) นำคะแนนสอบของนักเรียนทั้ง 11 คน มาเขียนเรียงจากน้อยไปมาก พร้อมทั้งแสดง Q_1, Q_2 และ Q_3 ได้ดังนี้

3 5 **67** 68 70 **70** 71 72 **73** 74 78

67 ↑ Q_1 70 ↑ Q_2 73 ↑ Q_3

แทน Q_1 และ Q_3 ด้วย 67 และ 73 ตามลำดับ ใน $Q_1 - 1.5(Q_3 - Q_1)$ จะได้

$67 - 1.5(73 - 67) = 58$

แทน Q_1 และ Q_3 ด้วย 67 และ 73 ตามลำดับ ใน $Q_3 + 1.5(Q_3 - Q_1)$ จะได้

$73 + 1.5(73 - 67) = 82$

จากข้อมูลข้างต้น มี 3 และ 5 น้อยกว่า 58 แต่ไม่มีข้อมูลที่มีค่ามากกว่า 82

ดังนั้น ค่านอกเกณฑ์ของข้อมูลชุดนี้ คือ 3 และ 5

2) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้คือ

$$\frac{70+72+68+3+71+74+70+67+73+5+78}{11} = \frac{651}{11} \approx 59.18$$

3) ค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มมาโดยไม่รวมค่านอกเกณฑ์ คือ

$$\frac{70+72+68+71+74+70+67+73+78}{9} = \frac{643}{9}$$
$$\approx 71.44$$

จากตัวอย่างข้างต้น จะเห็นว่าค่านอกเกณฑ์มีผลต่อค่าเฉลี่ยเลขคณิต เนื่องจากค่าเฉลี่ยเลขคณิตคำนวณจากข้อมูลทั้งหมด ดังนั้น ถ้าข้อมูลชุดใดมีค่านอกเกณฑ์ ค่าเฉลี่ยเลขคณิตอาจไม่สามารถนำมาใช้เป็นตัวแทนที่ดีของชุดข้อมูลนั้น แต่ในกรณีที่ชุดข้อมูลไม่มีค่านอกเกณฑ์ ค่าเฉลี่ยเลขคณิตอาจเป็นค่ากลางที่ใช้เป็นตัวแทนของชุดข้อมูลได้

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก

ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก (weighted arithmetic mean) เหมาะสำหรับใช้ในกรณีที่ข้อมูลแต่ละค่ามีความสำคัญไม่เท่ากัน เช่น การหาค่าเฉลี่ยเลขคณิตของคะแนนสอบ 4 วิชา ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน ถ้าใช้วิธีการหาค่าเฉลี่ยเลขคณิตโดยไม่ถ่วงน้ำหนัก อาจทำให้ค่าที่ได้นำไปสู่ข้อสรุปที่คลาดเคลื่อนจากความเป็นจริง เพราะข้อมูลแต่ละค่ามีความสำคัญไม่เท่ากัน ขึ้นอยู่กับน้ำหนักของแต่ละข้อมูล

ให้ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร
และให้ $w_1, w_2, w_3, \dots, w_N$ แทนน้ำหนักของข้อมูล $x_1, x_2, x_3, \dots, x_N$ ตามลำดับ จะได้

$$\textbf{ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก} = \frac{w_1x_1 + w_2x_2 + w_3x_3 + \cdots + w_Nx_N}{w_1 + w_2 + w_3 + \cdots + w_N}$$
$$= \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$$

หมายเหตุ ในกรณีที่เป็นข้อมูลของตัวอย่าง สามารถหาค่าเฉลี่ยเลขคณิตถ่วงน้ำหนักได้ในทำนองเดียวกันกับสูตรข้างต้น โดยเปลี่ยน N เป็น n เมื่อ N แทนขนาดประชากร และ n แทนขนาดตัวอย่าง

ตัวอย่างที่ 12

ในการคำนวณเกรดเฉลี่ย (grade point average: GPA) ของนักเรียนคนหนึ่ง สมมติว่านักเรียนคนนี้ลงทะเบียนเรียน 5 วิชา ซึ่งแต่ละวิชามีหน่วยกิตไม่เท่ากัน และได้เกรดแต่ละวิชาดังนี้

วิชา	คณิตศาสตร์	ภาษาอังกฤษ	ภาษาไทย	สังคมศึกษา	สุขศึกษา
หน่วยกิต	2.0	3.0	3.0	3.0	1.0
เกรด	4	4	3	3	2

จงหาเกรดเฉลี่ยของนักเรียนคนนี้

วิธีทำ ในที่นี้ข้อมูลคือเกรดแต่ละวิชาของนักเรียนคนนี้ ซึ่งได้แก่ 4, 4, 3, 3, 2 และหน่วยกิต 2.0, 3.0, 3.0, 3.0, 1.0 คือน้ำหนักของข้อมูล 4, 4, 3, 3, 2 ตามลำดับ

จะได้ ค่าเฉลี่ยเลขคณิตถ่วงน้ำหนัก
$$= \frac{2.0(4)+3.0(4)+3.0(3)+3.0(3)+1.0(2)}{2.0+3.0+3.0+3.0+1.0}$$
$$= \frac{40}{12}$$
$$\approx 3.33$$

ดังนั้น เกรดเฉลี่ยของนักเรียนคนนี้ประมาณ 3.33 ■

มัธยฐาน

เมื่อนำข้อมูลทั้งหมดมาเรียงลำดับจากน้อยไปมากหรือจากมากไปน้อย จะเรียกค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลว่า **มัธยฐาน (median)**

ถ้าข้อมูลมี n ตัว การหามัธยฐานทำได้โดยเรียงลำดับข้อมูลจากน้อยไปมากหรือจากมากไปน้อย จะได้มัธยฐานอยู่ในตำแหน่งที่ $\frac{n+1}{2}$ นั่นคือ

- ถ้า n เป็นจำนวนคี่ มัธยฐานคือข้อมูลที่อยู่กึ่งกลาง
- ถ้า n เป็นจำนวนคู่ มัธยฐานคือค่าเฉลี่ยเลขคณิตของข้อมูลสองตัวที่อยู่กึ่งกลาง

จากหัวข้อ 3.2 นักเรียนทราบแล้วว่า Q_2 เป็นค่าที่อยู่ในตำแหน่งกึ่งกลางของข้อมูลทั้งหมด ดังนั้น มัธยฐาน คือ Q_2

ตัวอย่างที่ 13

ความสูง (เซนติเมตร) ของนักเรียนหญิงจำนวน 11 คน แสดงได้ดังนี้

164 158 167 160 163 159 162 161 155 170 168

จงหามัธยฐานของข้อมูลชุดนี้

วิธีทำ เรียงความสูงของนักเรียนหญิง 11 คน จากน้อยไปมาก ได้ดังนี้

155 158 159 160 161 162 163 164 167 168 170

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $\frac{11+1}{2} = 6$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ 162 เซนติเมตร ■

ตัวอย่างที่ 14

ระยะเวลา (นาที) ที่ใช้ในการเดินทางจากบ้านไปโรงเรียนของนักเรียน 6 คน แสดงได้ดังนี้

32 15 45 12 90 25

จงหามัธยฐานของข้อมูลชุดนี้

วิธีทำ เรียงระยะเวลาที่ใช้ในการเดินทางจากบ้านไปโรงเรียนของนักเรียน 6 คน จากน้อยไปมาก ได้ดังนี้

12 15 25 32 45 90

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $\frac{6+1}{2} = 3.5$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ ค่าเฉลี่ยเลขคณิตของข้อมูลในตำแหน่งที่ 3 และ 4

ซึ่งคือ $\frac{25+32}{2} = 28.5$ นาที ■

จากตัวอย่างข้างต้น จะเห็นว่า 90 เป็นค่าที่สูงกว่าค่าส่วนใหญ่ของข้อมูลชุดนี้ แต่ไม่มีผลต่อลำดับของข้อมูล ดังนั้น 90 จึงไม่มีผลต่อมัธยฐานของข้อมูลชุดนี้

ฐานนิยม

จากบทที่ 2 นักเรียนได้ศึกษามาแล้วว่าสำหรับข้อมูลเชิงคุณภาพ ฐานนิยม คือ ข้อมูลที่มีจำนวนครั้งของการเกิดซ้ำกันมากที่สุดหรือข้อมูลที่มีความถี่สูงสุดที่มากกว่า 1 ซึ่งบทนิยามของฐานนิยมดังกล่าวสามารถใช้กับข้อมูลเชิงปริมาณได้เช่นกัน

ข้อมูลบางชุดอาจไม่มีฐานนิยม เช่น ในกรณีที่ข้อมูลมีความถี่เป็น 1 เท่ากันหมด นอกจากนี้ ข้อมูลบางชุดอาจมีฐานนิยมมากกว่า 1 ค่า เช่น ในกรณีที่มีข้อมูลมากกว่า 1 ข้อมูล ที่มีความถี่สูงสุดเท่ากัน

อย่างไรก็ตาม ในที่นี้จะพิจารณาเฉพาะชุดข้อมูลที่มีฐานนิยมเพียงค่าเดียว

ตัวอย่างที่ 15

อายุ (ปี) ของนักเรียนที่มาเข้าค่ายคณิตศาสตร์ จำนวน 15 คน แสดงได้ดังนี้

5	8	7	6	7	8	12	11
10	11	8	6	8	7	8	

จงหาฐานนิยมของข้อมูลชุดนี้

วิธีทำ จากข้อมูลที่กำหนดให้ จะได้

อายุ 5 ปี มีความถี่เป็น 1
อายุ 6 ปี มีความถี่เป็น 2
อายุ 7 ปี มีความถี่เป็น 3
อายุ 8 ปี มีความถี่เป็น 5
อายุ 10 ปี มีความถี่เป็น 1
อายุ 11 ปี มีความถี่เป็น 2
อายุ 12 ปี มีความถี่เป็น 1

จะเห็นว่า อายุ 8 ปี มีความถี่สูงสุด
ดังนั้น ฐานนิยมของข้อมูลชุดนี้ คือ 8 ปี ■

ข้อสังเกตที่สำคัญเกี่ยวกับค่ากลางชนิดต่าง ๆ

1. ฐานนิยมจะมีค่าตรงกับค่าใดค่าหนึ่งของข้อมูลชุดนั้น ในขณะที่ค่าเฉลี่ยเลขคณิตและมัธยฐานอาจไม่ใช่ค่าใดค่าหนึ่งของข้อมูลชุดนั้น
2. โดยปกติค่าเฉลี่ยเลขคณิตมักเป็นค่ากลางที่นิยมมากที่สุด แต่ถ้าชุดข้อมูลมีข้อมูลที่แตกต่างจากข้อมูลตัวอื่นมากจะมีผลต่อค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ แต่จะไม่มีผลต่อมัธยฐานและฐานนิยม
3. สำหรับข้อมูลเชิงคุณภาพ จะสามารถหาค่ากลางได้เฉพาะฐานนิยมเท่านั้น ไม่สามารถหาค่าเฉลี่ยเลขคณิตและมัธยฐานได้
4. ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมไม่จำเป็นต้องมีค่าเท่ากัน ทั้งนี้ ค่าเฉลี่ยเลขคณิตและมัธยฐานจะมีค่าที่ไม่สูงหรือต่ำเกินไปเมื่อเทียบกับค่าของข้อมูลทั้งหมด ในขณะที่ฐานนิยมอาจเป็นค่าสูงสุดหรือค่าต่ำสุดของชุดข้อมูลนั้นได้

การพิจารณาเลือกใช้ค่ากลางของข้อมูลไม่มีกฎเกณฑ์ที่แน่ชัด แต่ควรเลือกให้เหมาะสมกับวัตถุประสงค์ในการนำไปใช้และลักษณะของข้อมูล รวมทั้งต้องพิจารณาข้อดีและข้อเสียของค่ากลางแต่ละชนิด หากเลือกใช้ค่ากลางที่ไม่เหมาะสม อาจทำให้สรุปผลหรือตัดสินใจผิดพลาดได้

ตัวอย่างที่ 16

เงินเดือน (บาท) ของพนักงานแผนกหนึ่งในบริษัทแห่งหนึ่งจำนวนทั้งหมด 7 คน แสดงได้ดังนี้

15,300 16,600 13,450 15,300 14,400 15,300 71,000

จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลชุดนี้ และพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

วิธีทำ ให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

จะได้
$$\mu = \frac{15,300+16,600+13,450+15,300+14,400+15,300+71,000}{7}$$
$$= \frac{161,350}{7}$$
$$= 23,050$$

ดังนั้น ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ คือ 23,050 บาท

เรียงเงินเดือนของพนักงานทั้ง 7 คน จากน้อยไปมาก ได้ดังนี้

13,450 14,400 15,300 15,300 15,300 16,600 71,000

เนื่องจากมัธยฐานอยู่ในตำแหน่งที่ $\frac{7+1}{2} = 4$

ดังนั้น มัธยฐานของข้อมูลชุดนี้ คือ 15,300 บาท

จากข้อมูลที่กำหนดให้ จะเห็นว่า เงินเดือน 15,300 บาท มีความถี่สูงสุด

ดังนั้น ฐานนิยมของข้อมูลชุดนี้ คือ 15,300 บาท

สังเกตว่า 71,000 เป็นค่าที่สูงกว่าค่าส่วนใหญ่ของข้อมูลชุดนี้ หากใช้ค่าเฉลี่ยเลขคณิตเป็นตัวแทนของข้อมูลชุดนี้ จะเห็นได้ว่ามีพนักงานจำนวน 6 จาก 7 คน ที่ได้รับเงินเดือนน้อยกว่าค่าเฉลี่ยเลขคณิต ดังนั้น ค่าเฉลี่ยเลขคณิตจึงไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้ อย่างไรก็ตาม 71,000 ไม่มีผลต่อมัธยฐานและฐานนิยม และจากตัวอย่างนี้จะเห็นว่าทั้งสองค่านี้เท่ากัน

ดังนั้น ควรใช้มัธยฐานและฐานนิยมเป็นตัวแทนของข้อมูลชุดนี้ ■

ตัวอย่างที่ 17

นักเรียนคนหนึ่งได้คะแนนสอบย่อยวิชาคณิตศาสตร์จำนวน 5 ครั้ง ซึ่งแต่ละครั้งมีคะแนนเต็มเท่ากัน ดังนี้

17 17 17 19 20

จงพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

วิธีทำ จากข้อมูลที่กำหนดให้

จะได้ ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ คือ $\frac{17+17+17+19+20}{5}=18$ คะแนน

มัธยฐานของข้อมูลชุดนี้ คือ 17 คะแนน

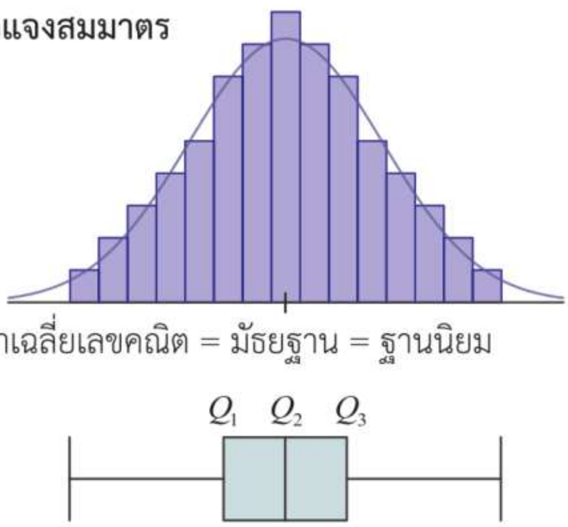
และ ฐานนิยมของข้อมูลชุดนี้ คือ 17 คะแนน

จะเห็นว่าในสถานการณ์นี้ มัธยฐานและฐานนิยมไม่เหมาะสมที่จะเป็นตัวแทนของข้อมูลชุดนี้ เนื่องจากไม่ได้ใช้คะแนนสอบย่อยทุกครั้งในการคำนวณ ในขณะที่ค่าเฉลี่ยเลขคณิตได้จากการนำคะแนนสอบย่อยทุกครั้งมาคำนวณ และคะแนนสอบย่อยแต่ละครั้งใกล้เคียงกัน ดังนั้น ควรใช้ค่าเฉลี่ยเลขคณิตเป็นตัวแทนของข้อมูลชุดนี้ ■

ความสัมพันธ์ระหว่างการกระจายของข้อมูลและค่ากลางของข้อมูล

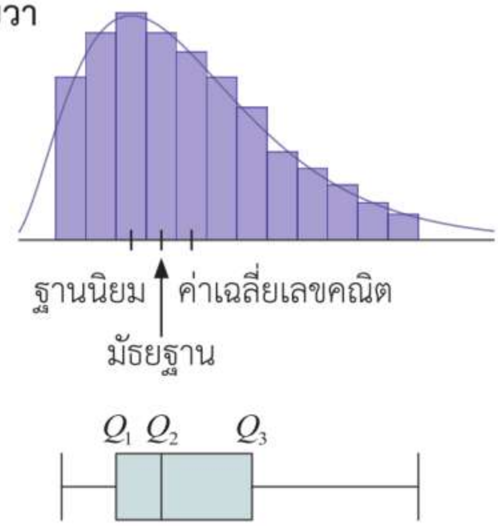
การอธิบายลักษณะการกระจายของข้อมูล นอกจากจะวิเคราะห์โดยใช้แผนภาพกล่องตามที่ได้ศึกษาในหัวข้อ 3.2 แล้ว ยังสามารถวิเคราะห์ได้โดยใช้ความสัมพันธ์ของค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยม ในที่นี้จะแบ่งลักษณะการกระจายของข้อมูลเป็น 3 แบบ ดังนี้

การแจกแจงสมมาตร



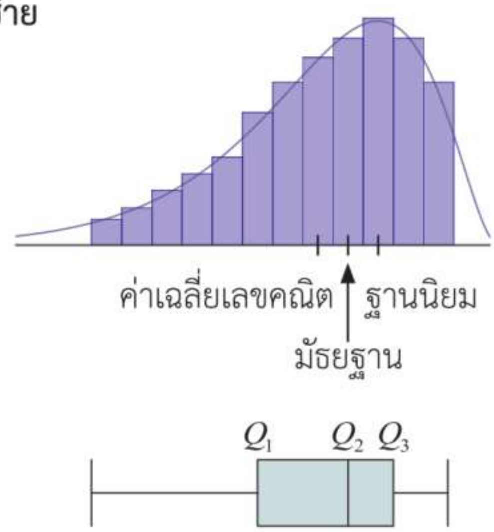
รูปที่ 4

การแจกแจงเบ้ขวา



รูปที่ 5

การแจกแจงเบ้ซ้าย



รูปที่ 6

ลักษณะการกระจายของข้อมูลในรูปที่ 4 เรียกว่า **การแจกแจงสมมาตร (symmetrical distribution)** และจากรูปที่ 4 จะได้ความสัมพันธ์ของค่ากลางของข้อมูล ดังนี้

ค่าเฉลี่ยเลขคณิต = มัธยฐาน = ฐานนิยม

จะเห็นว่าข้อมูลที่มีความถี่สูงสุดจะอยู่ตรงกลางและความถี่ของข้อมูลจะลดลงเมื่อข้อมูลมีค่าห่างจากมัธยฐาน เมื่อพิจารณาจากแผนภาพกล่องจะเห็นว่าความกว้างของช่วงจาก Q_1 ถึง Q_2 เท่ากับความกว้างของช่วงจาก Q_2 ถึง Q_3

ลักษณะการกระจายของข้อมูลในรูปที่ 5 เรียกว่า **การแจกแจงเบ้ขวา (right-skewed distribution)** โดยมีความสัมพันธ์ของค่ากลางของข้อมูลดังนี้

ฐานนิยม < มัธยฐาน < ค่าเฉลี่ยเลขคณิต

จะเห็นว่าข้อมูลที่มีค่าน้อยจะมีความถี่สูง และความถี่ของข้อมูลจะลดลงเมื่อค่าของข้อมูลเพิ่มขึ้น เมื่อพิจารณาจากแผนภาพกล่องจะเห็นว่าความกว้างของช่วงจาก Q_1 ถึง Q_2 น้อยกว่าความกว้างของช่วงจาก Q_2 ถึง Q_3

ลักษณะการกระจายของข้อมูลในรูปที่ 6 เรียกว่า **การแจกแจงเบ้ซ้าย (left-skewed distribution)** โดยมีความสัมพันธ์ของค่ากลางของข้อมูลดังนี้

ค่าเฉลี่ยเลขคณิต < มัธยฐาน < ฐานนิยม

จะเห็นว่าข้อมูลที่มีค่ามากจะมีความถี่สูง และความถี่ของข้อมูลจะลดลงเมื่อค่าของข้อมูลลดลง เมื่อพิจารณาจากแผนภาพกล่องจะเห็นว่าความกว้างของช่วงจาก Q_1 ถึง Q_2 มากกว่าความกว้างของช่วงจาก Q_2 ถึง Q_3

ถึงแม้ค่ากลางของข้อมูลจะสามารถใช้ในการบอกลักษณะการกระจายของข้อมูล แต่ก็ยังไม่สามารถบอกได้ว่าข้อมูลมีการกระจายมากหรือน้อย ในหัวข้อต่อไปนักเรียนจะได้ศึกษาค่าที่ใช้ในการพิจารณาว่าข้อมูลมีการกระจายมากหรือน้อยเพียงใด

แบบฝึกหัด 3.3.1

1. จากการสุ่มเก็บข้อมูลค่าจ้างรายวัน (บาท) ของพนักงานชั่วคราวของร้านสะดวกซื้อ 2 แห่ง เป็นเวลา 10 วัน ในเดือนมกราคม พ.ศ. 2562 แสดงได้ดังนี้

ร้านที่ 1	248	225	280	324	346	320	284	275	325	375
ร้านที่ 2	260	232	245	220	256	248	276	235	244	280

1) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของค่าจ้างรายวันของพนักงานแต่ละร้าน
2) นักเรียนจะเลือกทำงานที่ร้านใด เพราะเหตุใด

2. ธนาคารแห่งหนึ่งสำรวจระยะเวลา (นาที) ที่ลูกค้าใช้ในการทำธุรกรรมที่ธนาคาร ได้ผลสำรวจดังนี้

14	13	17	15	15	14	15	28	18
17	11	9	13	16	18	15	14	16
7	16	11	12	19	27	14	12	19

จงหา
1) ค่านอกเกณฑ์ของข้อมูลชุดนี้
2) ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้
3) ค่าเฉลี่ยเลขคณิตของระยะเวลาที่ลูกค้าใช้ในการทำธุรกรรมที่ธนาคารแห่งนี้ โดยไม่รวมค่านอกเกณฑ์
4) ค่าเฉลี่ยเลขคณิตที่ได้จากข้อ 2) และ 3) แตกต่างกันมากหรือไม่ เพราะเหตุใด

3. วิชาคณิตศาสตร์มีการสอบ 3 ครั้ง เป็นการสอบย่อย 2 ครั้ง และสอบปลายภาค 1 ครั้ง โดยคะแนนสอบย่อยแต่ละครั้งคิดเป็นร้อยละ 15 ของคะแนนทั้งหมด และคะแนนสอบปลายภาคคิดเป็นร้อยละ 70 ของคะแนนทั้งหมด ถ้านักเรียนคนหนึ่งได้คะแนนสอบย่อย 2 ครั้ง เป็น 74 และ 80 คะแนน และได้คะแนนสอบปลายภาค 62 คะแนน โดยแต่ละครั้งมีคะแนนเต็ม 100 คะแนน จงหาคะแนนเฉลี่ยวิชาคณิตศาสตร์ของนักเรียนคนนี้

4. นักสัตววิทยาคนหนึ่งสำรวจจำนวนการตายและการเกิดของไก่ป่าชนิดหนึ่งในพื้นที่สำรวจ 14 พื้นที่ ในเดือนพฤษภาคม พ.ศ. 2562 ได้ผลสำรวจดังนี้

พื้นที่สำรวจ	จำนวนการตายของไก่ป่า (ตัว)	จำนวนการเกิดของไก่ป่า (ตัว)
1	5	30
2	0	28
3	0	38
4	8	34
5	9	26
6	7	40
7	2	48
8	6	46
9	4	32
10	0	31
11	2	46
12	10	132
13	3	42
14	5	126

1) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนการตายของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ และพิจารณาว่าค่ากลางใดไม่ควรเป็นตัวแทนของจำนวนการตายของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ พร้อมทั้งให้เหตุผลประกอบ
2) จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนการเกิดของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ และพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของจำนวนการเกิดของไก่ป่าในพื้นที่สำรวจทั้ง 14 พื้นที่ พร้อมทั้งให้เหตุผลประกอบ

5. ในการสมัครเข้ารับการคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัยแห่งหนึ่ง ผู้สมัครต้องยื่นจดหมายรับรองจำนวน 3 ฉบับ เพื่อประกอบการพิจารณา ถ้าค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของจำนวนจดหมายรับรองของผู้สมัครจำนวน 148 คน คือ 2.9, 3 และ 3 ฉบับ ตามลำดับ จงใช้ค่ากลางดังกล่าวในการอธิบายความหมายของจำนวนจดหมายรับรองของผู้สมัครทั้ง 148 คน

6. ถ้าค่าเฉลี่ยเลขคณิตของน้ำหนักของนักเรียน 3 คน คือ 38 กิโลกรัม และนักเรียนหนึ่งคนในกลุ่มนี้หนัก 46 กิโลกรัม ส่วนอีกสองคนที่เหลือหนักเท่ากัน จงหาว่านักเรียนสองคนที่เหลือหนักคนละกี่กิโลกรัม

7. ข้อมูลชุดหนึ่งมี 7 ตัว และมีค่าเฉลี่ยเลขคณิตคือ 81 ถ้าตัดข้อมูลออกไป 1 ตัว แล้วทำให้ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เหลือ 78 จงหาว่าข้อมูลที่ถูกตัดออกไปมีค่าเท่าใด

8. วิชาคณิตศาสตร์มีการสอบย่อยทั้งหมด 5 ครั้ง แต่ละครั้งมีคะแนนเต็มเท่ากัน ถ้าค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของคะแนนสอบย่อยทั้งห้าครั้งของน้อยหน่า คือ 86, 87 และ 80 คะแนน ตามลำดับ จงหาคะแนนสอบย่อยที่สูงที่สุดที่เป็นไปได้ของน้อยหน่า ถ้าคะแนนสอบทั้งห้าครั้งของน้อยหน่าเป็นจำนวนเต็ม

3.3.2 ค่าวัดการกระจาย

การทราบเพียงค่ากลางของข้อมูลไม่เพียงพอที่จะบอกว่าข้อมูลมีการกระจายมากหรือน้อย เนื่องจากค่ากลางแต่ละชนิดมิได้บอกให้ทราบว่าข้อมูลแต่ละค่าห่างกันมากหรือน้อยเพียงใด ข้อมูลส่วนใหญ่รวมกลุ่มกันหรือกระจายกันออกไป สมมติว่าในการสอบวิชาคณิตศาสตร์ของนักเรียนสองห้องซึ่งใช้ข้อสอบชุดเดียวกันมีค่าเฉลี่ยเลขคณิตของคะแนนสอบเท่ากันคือ 67 คะแนน ห้องแรกมีคะแนนสูงสุด 72 คะแนน และคะแนนต่ำสุด 62 คะแนน ส่วนห้องที่สองมีคะแนนสูงสุด 97 คะแนน และคะแนนต่ำสุด 25 คะแนน จะเห็นว่าคะแนนสูงสุดและคะแนนต่ำสุดของห้องแรกต่างกันเพียง 10 คะแนน แต่คะแนนสูงสุดและคะแนนต่ำสุดของห้องที่สองต่างกันถึง 72 คะแนน แสดงว่าคะแนนของห้องที่สองมีการกระจายมากกว่าห้องแรกมาก ซึ่งอาจกล่าวได้ว่านักเรียนห้องแรกส่วนใหญ่สอบได้คะแนนใกล้เคียงกัน แต่นักเรียนห้องที่สองสอบได้คะแนนแตกต่างกันมาก ต่อไปนักเรียนจะได้ศึกษาค่าวัดทางสถิติที่จะช่วยให้เห็นลักษณะของข้อมูลชัดเจนขึ้นและสามารถวิเคราะห์เกี่ยวกับข้อมูลนั้นได้มากขึ้น

โดยทั่วไป การวัดการกระจายของข้อมูลแบ่งได้เป็น 2 วิธี คือ

1. **การกระจายสัมบูรณ์ (absolute variation)** คือ การวัดการกระจายของข้อมูลด้วยค่าวัดทางสถิติที่มีหน่วยเช่นเดียวกับข้อมูลหรือเป็นกำลังสองของหน่วยของข้อมูล เพื่อใช้พิจารณาว่าข้อมูลแต่ละตัวมีความแตกต่างกันมากหรือน้อยเพียงใด ในที่นี้จะศึกษาค่าวัดการกระจายสัมบูรณ์ 4 ชนิด คือ
   1) พิสัย
   2) พิสัยระหว่างควอร์ไทล์
   3) ส่วนเบี่ยงเบนมาตรฐาน
   4) ความแปรปรวน

2. **การกระจายสัมพัทธ์ (relative variation)** คือ การวัดการกระจายของข้อมูลด้วยค่าวัดทางสถิติที่ไม่มีหน่วย ซึ่งเป็นค่าที่ใช้ในการเปรียบเทียบการกระจายระหว่างข้อมูลมากกว่า 1 ชุด ในที่นี้จะศึกษาค่าวัดการกระจายสัมพัทธ์เพียงชนิดเดียวคือสัมประสิทธิ์การแปรผัน

การวัดการกระจายสัมบูรณ์

1) พิสัย

พิสัย (range) คือ ค่าที่ใช้วัดการกระจายของข้อมูลชุดหนึ่ง โดยคำนวณจากผลต่างระหว่างค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดนั้น

กำหนดให้ข้อมูลชุดหนึ่งมี x_{max} และ x_{min} เป็นค่าสูงสุดและค่าต่ำสุด ตามลำดับ

$$\textbf{พิสัย} = x_{\text{max}} - x_{\text{min}}$$

ตัวอย่างที่ 18

ผลผลิตน้ำตาลใน พ.ศ. 2561/62 ของจีน สหรัฐอเมริกา ไทย อินเดีย ออสเตรเลีย และบราซิล แสดงได้ดังนี้

ประเทศ	จีน	สหรัฐอเมริกา	ไทย	อินเดีย	ออสเตรเลีย	บราซิล
ผลผลิต (ล้านตัน)	10.60	8.12	14.19	33.07	4.90	29.50

จงหาพิสัยของข้อมูลชุดนี้

วิธีทำ จากโจทย์ จะได้ว่าค่าสูงสุดและค่าต่ำสุดของข้อมูลชุดนี้ คือ 33.07 และ 4.90 ล้านตัน ตามลำดับ

ดังนั้น พิสัยของข้อมูลชุดนี้ คือ 33.07 – 4.90 = 28.17 ล้านตัน ■

จากตัวอย่างข้างต้น สามารถนำพิสัยมาใช้ในการอธิบายอย่างคร่าว ๆ ว่าผลผลิตน้ำตาลใน พ.ศ. 2561/62 ของแต่ละประเทศจะต่างกันไม่เกิน 28.17 ล้านตัน

เสริมสมอง : ผลผลิตน้ำตาลใน พ.ศ. 2561/62

ผลผลิตน้ำตาลใน พ.ศ. 2561/62 หมายถึงผลผลิตน้ำตาลในปีการตลาด 2561/62 โดยปีการตลาดของจีน สหรัฐอเมริกา ไทย อินเดีย ออสเตรเลีย และบราซิล พิจารณาจากช่วงเวลาดังต่อไปนี้

ประเทศ	ปีการตลาด 2561/62
จีน	ตุลาคม พ.ศ. 2561 – กันยายน พ.ศ. 2562
สหรัฐอเมริกา	ตุลาคม พ.ศ. 2561 – กันยายน พ.ศ. 2562
ไทย	ธันวาคม พ.ศ. 2561 – พฤศจิกายน พ.ศ. 2562
อินเดีย	ตุลาคม พ.ศ. 2561 – กันยายน พ.ศ. 2562
ออสเตรเลีย	กรกฎาคม พ.ศ. 2561 – มิถุนายน พ.ศ. 2562
บราซิล	เมษายน พ.ศ. 2561 – มีนาคม พ.ศ. 2562

ข้อดีของการใช้พิสัยในการวัดการกระจายของข้อมูลคือสามารถหาได้สะดวก แต่การวัดการกระจายของข้อมูลโดยใช้พิสัยเป็นการวัดการกระจายของข้อมูลอย่างคร่าว ๆ เพราะพิสัยคำนวณจากข้อมูลเพียงสองค่าเท่านั้นคือค่าสูงสุดและค่าต่ำสุด ไม่ได้ใช้ข้อมูลอื่น ๆ ในการคำนวณเลย ดังนั้น การใช้พิสัยในการวัดการกระจายของข้อมูลอาจให้ข้อสรุปที่คลาดเคลื่อน ในกรณีที่ชุดข้อมูลมีข้อมูลที่มีค่าสูงหรือต่ำกว่าข้อมูลตัวอื่นมาก เช่น คะแนนสอบวิชาภาษาอังกฤษของนักเรียนจำนวน 10 คน เป็นดังนี้

10 70 71 72 73 74 75 76 77 100

จะเห็นว่านักเรียนส่วนใหญ่ได้คะแนนใกล้เคียงกัน โดยมีค่าตั้งแต่ 70 ถึง 77 คะแนน ยกเว้นนักเรียนที่ได้คะแนนสูงสุดและต่ำสุด แต่พิสัยของข้อมูลชุดนี้ คือ $100 - 10 = 90$ คะแนน ทำให้อาจเข้าใจว่านักเรียนได้คะแนนแตกต่างกันมาก ซึ่งคลาดเคลื่อนไปจากความเป็นจริง

2) พิสัยระหว่างควอร์ไทล์

พิสัยระหว่างควอร์ไทล์ (interquartile range) คือ ค่าที่ใช้วัดการกระจายของข้อมูลโดยคำนวณจากผลต่างระหว่างควอร์ไทล์ที่สามและควอร์ไทล์ที่หนึ่ง เขียนแทนพิสัยระหว่างควอร์ไทล์ด้วย IQR

ให้ Q_1 และ Q_3 เป็นควอร์ไทล์ที่หนึ่งและควอร์ไทล์ที่สามของข้อมูลชุดหนึ่งตามลำดับ จะได้

$$\text{IQR} = Q_3 - Q_1$$

การวัดการกระจายสัมบูรณ์โดยใช้พิสัยระหว่างควอร์ไทล์มีข้อดีในกรณีที่ชุดข้อมูลมีข้อมูลที่แตกต่างจากข้อมูลตัวอื่นมาก เนื่องจากการคำนวณหาพิสัยระหว่างควอร์ไทล์จะใช้เพียง Q_1 และ Q_3 เท่านั้น ส่วนข้อมูลที่แตกต่างจากข้อมูลตัวอื่นมากจะมีค่าน้อยกว่า Q_1 หรือมากกว่า Q_3

นอกจากนี้ IQR สามารถนำไปใช้ในการตรวจสอบว่าข้อมูลใดเป็นค่านอกเกณฑ์ ดังที่ได้นำเสนอไว้ในหัวข้อ 3.2 เรื่องแผนภาพกล่อง นั่นคือ ค่านอกเกณฑ์คือข้อมูลที่มีค่าน้อยกว่า $Q_1 - 1.5\text{IQR}$ หรือมากกว่า $Q_3 + 1.5\text{IQR}$

ตัวอย่างที่ 19

ปริมาณพลังงาน (กิโลแคลอรี) ของอาหารจานเดียว 7 รายการ ที่จำหน่ายในโรงอาหารของโรงเรียนแห่งหนึ่ง แสดงได้ดังนี้

อาหารจานเดียว	ปริมาณพลังงาน (กิโลแคลอรี)
ข้าวราดแกงเขียวหวานไก่	338
ข้าวราดแกงไตปลา	319
ข้าวราดแกงส้มผักรวม	255
ข้าวราดผัดเผ็ดหอยลาย	424
ข้าวราดแกงพะแนงหมู	409
ข้าวราดแกงฉู่ฉี่ปลาทู	365
ข้าวราดผัดผักรวม	353

จงหาพิสัยระหว่างควอร์ไทล์ของข้อมูลชุดนี้

วิธีทำ จากข้อมูลข้างต้น เรียงปริมาณพลังงานของอาหารจานเดียวทั้ง 7 รายการ จากน้อยไปมาก ได้ดังนี้

255 319 338 353 365 409 424

เนื่องจาก Q_1 อยู่ในตำแหน่งที่ $\frac{1(7+1)}{4} = 2$

จะได้ $Q_1 = 319$

เนื่องจาก Q_3 อยู่ในตำแหน่งที่ $\frac{3(7+1)}{4} = 6$

จะได้ $Q_3 = 409$

ดังนั้น พิสัยระหว่างควอร์ไทล์ของข้อมูลชุดนี้ คือ $409 - 319 = 90$ กิโลแคลอรี

เมื่อเปรียบเทียบระหว่างพิสัยและพิสัยระหว่างควอร์ไทล์ จะเห็นว่าพิสัยสามารถหาได้สะดวก แต่ไม่เหมาะสำหรับใช้วัดการกระจายของข้อมูลในกรณีที่ชุดข้อมูลมีข้อมูลที่มีค่าสูงหรือต่ำกว่าข้อมูลตัวอื่นมาก โดยเฉพาะอย่างยิ่งในกรณีที่ชุดข้อมูลมีค่านอกเกณฑ์ ในขณะที่พิสัยระหว่างควอร์ไทล์สามารถใช้วัดการกระจายของข้อมูลในลักษณะนี้ได้ อย่างไรก็ตาม ทั้งพิสัยและพิสัยระหว่างควอร์ไทล์ไม่ได้ใช้ข้อมูลทุกตัวในการคำนวณเพื่อวัดการกระจาย

3) ส่วนเบี่ยงเบนมาตรฐาน

ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) คือ ค่าที่ใช้วัดการกระจายของข้อมูล โดยเป็นค่าที่บอกให้ทราบว่าข้อมูลแต่ละตัวอยู่ห่างจากค่าเฉลี่ยเลขคณิตโดยเฉลี่ยประมาณเท่าใด สูตรของส่วนเบี่ยงเบนมาตรฐานมีดังนี้

ให้ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร และให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ส่วนเบี่ยงเบนมาตรฐานของประชากร เขียนแทนด้วย σ (อ่านว่า ซิกมา) หาได้จาก

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$$

ให้ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง และให้ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง เขียนแทนด้วย s หาได้จาก

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

ตัวอย่างที่ 20

ความสูง (เซนติเมตร) ของนักวอลเลย์บอลหญิงของโรงเรียนแห่งหนึ่งจำนวนทั้งหมด 10 คน แสดงได้ดังนี้

174	171	170	184	180
179	169	178	181	160

จงหาส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

วิธีทำ ให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

และ x_i แทนความสูงของนักวอลเลย์บอลหญิงคนที่ i เมื่อ $i \in \{1, 2, 3, \dots, 10\}$

จะได้

$$\mu = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{1,746}{10} = 174.6$$

ดังนั้น ค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้ คือ 174.6 เซนติเมตร

จากข้อมูลข้างต้น จะได้

x_i	$x_i - \mu$	$(x_i - \mu)^2$
174	– 0.6	0.36
171	– 3.6	12.96
170	– 4.6	21.16
184	9.4	88.36
180	5.4	29.16
179	4.4	19.36
169	– 5.6	31.36
178	3.4	11.56
181	6.4	40.96
160	– 14.6	213.16
		$\sum_{i=1}^{10} (x_i - \mu)^2 = 468.40$

ดังนั้น $\sigma = \sqrt{\frac{468.40}{10}} \approx 6.84$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้มีค่าประมาณ 6.84 เซนติเมตร ■

จากตัวอย่างข้างต้น สามารถนำส่วนเบี่ยงเบนมาตรฐานมาใช้ในการอธิบายว่าโดยเฉลี่ยแล้ว ความสูงของนักวอลเลย์บอลหญิงแต่ละคนของโรงเรียนแห่งนี้ต่างจากความสูงเฉลี่ยประมาณ 6.84 เซนติเมตร

4) ความแปรปรวน

ความแปรปรวน (variance) คือ ค่าที่ใช้วัดการกระจายของข้อมูล โดยคำนวณจากกำลังสองของส่วนเบี่ยงเบนมาตรฐาน จะได้ สูตรของความแปรปรวน ดังนี้

ให้ $x_1, x_2, x_3, \dots, x_N$ แทนข้อมูล เมื่อ N แทนขนาดประชากร และให้ μ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ความแปรปรวนของประชากร หาได้จาก

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$$

ให้ $x_1, x_2, x_3, \dots, x_n$ แทนข้อมูล เมื่อ n แทนขนาดตัวอย่าง และให้ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้

ความแปรปรวนของตัวอย่าง หาได้จาก

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

จากตัวอย่างที่ 20 จะได้ว่าความแปรปรวนของความสูงของนักวอลเลย์บอลหญิงจำนวน 10 คน คือ $\frac{468.40}{10} = 46.84$ เซนติเมตร^2

ข้อสังเกต ส่วนเบี่ยงเบนมาตรฐานมีหน่วยเหมือนกับหน่วยของข้อมูล แต่ความแปรปรวนมีหน่วยเป็นกำลังสองของหน่วยของข้อมูล

ตัวอย่างที่ 21

ในการศึกษาอายุขัยเฉลี่ย (ปี) ของสัตว์เลี้ยงลูกด้วยน้ำนม นักวิทยาศาสตร์ได้สุ่มตัวอย่างสัตว์เลี้ยงลูกด้วยน้ำนมมา 10 ชนิด พบว่าอายุขัยเฉลี่ยของสัตว์แต่ละชนิดเป็นดังนี้

สัตว์เลี้ยงลูกด้วยน้ำนม	อายุขัยเฉลี่ย (ปี)
แมว	12
วัว	15
สุนัข	12
ลา	12
แพะ	8
หนูตะเภา	4
ม้า	20
หมู	10
กระต่าย	5
แกะ	12

จงหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้

วิธีทำ ให้ x_i แทนอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนมชนิดที่ i เมื่อ $i \in \{1, 2, 3, \ldots, 10\}$ และ $\bar{x}$ แทนค่าเฉลี่ยเลขคณิตของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้

จะได้

$$\bar{x} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{110}{10} = 11$$

ดังนั้น ค่าเฉลี่ยเลขคณิตของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ คือ 11 ปี จากข้อมูลข้างต้น จะได้

x_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$
12	1	1
15	4	16
12	1	1
12	1	1
8	– 3	9
4	– 7	49
20	9	81
10	–1	1
5	– 6	36
12	1	1
		$\sum_{i=1}^{10}(x_i - \overline{x})^2 = 196$

ดังนั้น $s = \sqrt{\frac{196}{10-1}} = \frac{14}{3} \approx 4.67$ และ $s^2 = \frac{196}{9} \approx 21.78$

นั่นคือ ส่วนเบี่ยงเบนมาตรฐานของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ มีค่าประมาณ 4.67 ปี

และความแปรปรวนของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้ มีค่าประมาณ 21.78 ปี^{2} ■

การวัดการกระจายสัมพัทธ์

ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไป เพื่อพิจารณาว่าข้อมูลชุดใดมีการกระจายมาก ข้อมูลชุดใดมีการกระจายน้อย ถ้านำค่าที่ได้จากการวัดการกระจายสัมบูรณ์ของข้อมูลแต่ละชุดมาเปรียบเทียบกันโดยตรง อาจให้ข้อสรุปที่คลาดเคลื่อนไปจากความเป็นจริง เช่น ข้อมูลชุดหนึ่งมีค่าตั้งแต่ 0 ถึง 10 มีส่วนเบี่ยงเบนมาตรฐาน 2.2 และข้อมูลอีกชุดหนึ่งมีค่าตั้งแต่ 200 ถึง 800 มีส่วนเบี่ยงเบนมาตรฐาน 60.5 ถ้าพิจารณาเฉพาะส่วนเบี่ยงเบนมาตรฐานของข้อมูลทั้งสองชุด อาจทำให้เข้าใจว่าข้อมูลชุดที่หนึ่งมีการกระจายน้อยกว่าข้อมูลชุดที่สองซึ่งอาจไม่ถูกต้องนัก เพราะค่าของข้อมูลสองชุดนี้ต่างกันมาก ค่ากลางและค่าวัดการกระจายของข้อมูลทั้งสองชุดย่อมต่างกันมากเช่นกัน เพื่อให้การเปรียบเทียบมีความหมาย จึงนิยมหาอัตราส่วนของค่าที่ได้จากการวัดการกระจายสัมบูรณ์กับค่ากลางของข้อมูลชุดนั้น ๆ แล้วจึงนำอัตราส่วนที่หาได้มาเปรียบเทียบกัน ในที่นี้จะพิจารณาเฉพาะ**สัมประสิทธิ์การแปรผัน (coefficient of variation)** โดยมีสูตรดังนี้

สัมประสิทธิ์การแปรผันของประชากร $= \dfrac{\sigma}{|\mu|}$ เมื่อ $\mu \neq 0$

สัมประสิทธิ์การแปรผันของตัวอย่าง $= \dfrac{s}{|\bar{x}|}$ เมื่อ $\bar{x} \neq 0$

สัมประสิทธิ์การแปรผันอาจเขียนในรูปเปอร์เซ็นต์ ได้ดังนี้

สัมประสิทธิ์การแปรผันของประชากร $= \dfrac{\sigma}{|\mu|} \times 100\%$ เมื่อ $\mu \neq 0$

สัมประสิทธิ์การแปรผันของตัวอย่าง $= \dfrac{s}{|\bar{x}|} \times 100\%$ เมื่อ $\bar{x} \neq 0$

เช่น ข้อมูลตัวอย่างชุดหนึ่งมี $s = 10$ และ $\bar{x} = 30$ จะได้ สัมประสิทธิ์การแปรผันของข้อมูลชุดนี้ คือ 0.33 หรือ 33%

การเปรียบเทียบการกระจายของข้อมูลโดยใช้สัมประสิทธิ์การแปรผันนั้น ถ้าสัมประสิทธิ์การแปรผันของข้อมูลชุดใดมีค่ามากกว่า หมายความว่าข้อมูลชุดนั้นมีการกระจายออกจากค่าเฉลี่ยเลขคณิตมากกว่า หรือกล่าวได้ว่าข้อมูลชุดนั้นเกาะกลุ่มกันน้อยกว่าข้อมูลอีกชุดหนึ่ง

ตัวอย่างที่ 22

ในการเปรียบเทียบคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนห้องหนึ่ง ซึ่งมีคะแนนเต็มวิชาละ 100 คะแนน ครูประจำชั้นได้สุ่มตัวอย่างนักเรียนห้องนี้มา 10 คน พบว่าคะแนนสอบแต่ละวิชาของนักเรียนแต่ละคนเป็นดังนี้

นักเรียนคนที่	1	2	3	4	5	6	7	8	9	10
คะแนนสอบวิชาคณิตศาสตร์	58	62	76	90	78	81	88	79	80	75
คะแนนสอบวิชาภาษาอังกฤษ	78	74	63	89	76	75	85	90	73	74

จงหาสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ พร้อมทั้งเปรียบเทียบการกระจายของคะแนนสอบทั้งสองวิชาของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้

วิธีทำ ให้ x_i และ y_i แทนคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนคนที่ i เมื่อ $i \in \{1, 2, 3, \ldots, 10\}$ ตามลำดับ

$\overline{x}$ และ $\overline{y}$ แทนค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ ตามลำดับ

s_x และ s_y แทนส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ ตามลำดับ

จะได้
$$\overline{x} = \frac{\sum_{i=1}^{10} x_i}{10} = 76.7$$

และ
$$\overline{y} = \frac{\sum_{i=1}^{10} y_i}{10} = 77.7$$

จากข้อมูลข้างต้น จะได้

x_i	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$
58	– 18.7	349.69
62	– 14.7	216.09
76	– 0.7	0.49
90	13.3	176.89
78	1.3	1.69
81	4.3	18.49
88	11.3	127.69
79	2.3	5.29
80	3.3	10.89
75	– 1.7	2.89
		$\sum_{i=1}^{10}\left(x_i - \overline{x}\right)^2 = 910.1$

และ

y_i	$y_i - \overline{y}$	$\left(y_i - \overline{y}\right)^2$
78	0.3	0.09
74	– 3.7	13.69
63	– 14.7	216.09
89	11.3	127.69
76	– 1.7	2.89
75	– 2.7	7.29
85	7.3	53.29
90	12.3	151.29
73	– 4.7	22.09
74	– 3.7	13.69
		$\sum_{i=1}^{10}\left(y_i - \overline{y}\right)^2 = 608.1$

ดังนั้น $s_x = \sqrt{\frac{910.1}{10-1}} \approx 10.06$ และ $s_y = \sqrt{\frac{608.1}{10-1}} \approx 8.22$

นั่นคือ สัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ คือ $\frac{10.06}{|76.7|} \approx 0.1312$

และสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ คือ $\frac{8.22}{|77.7|} \approx 0.1058$

เมื่อพิจารณาจากนักเรียนที่สุ่มตัวอย่างมา 10 คนนี้ จะเห็นว่าสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาคณิตศาสตร์มากกว่าสัมประสิทธิ์การแปรผันของคะแนนสอบวิชาภาษาอังกฤษ สรุปได้ว่าคะแนนสอบวิชาคณิตศาสตร์มีการกระจายมากกว่าคะแนนสอบวิชาภาษาอังกฤษ หรือกล่าวได้ว่าคะแนนสอบวิชาภาษาอังกฤษเกาะกลุ่มกันมากกว่าคะแนนสอบวิชาคณิตศาสตร์ ■

ข้อสังเกต สัมประสิทธิ์การแปรผันไม่มีหน่วย

แบบฝึกหัด 3.3.2

1. ร้านค้าจำหน่ายและรับติดตั้งประตูอัตโนมัติแห่งหนึ่งเก็บข้อมูลตัวอย่างเกี่ยวกับเวลา (นาที) ที่ใช้ในการติดตั้งประตูแต่ละบาน ได้ข้อมูลดังนี้

   28 32 24 46 44 40 54 38 32 42 36

   จงหาพิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนของเวลาที่ใช้ในการติดตั้งประตู

2. ปริมาณพลังงาน (กิโลแคลอรี) ของอาหารจานเดียว 11 รายการ ที่สุ่มตัวอย่างมาจากโรงอาหารแห่งหนึ่ง แสดงได้ดังนี้

อาหารจานเดียว	ปริมาณพลังงาน (กิโลแคลอรี)
หอยทอด	933
สุกี้น้ำรวมมิตร	117
ข้าวผัดหมู	553
ข้าวหมูแดง	444
ข้าวมันไก่	717
เส้นใหญ่ราดหน้าหมู	337
ข้าวหมกไก่	475
ข้าวคลุกกะปิ	522
หมี่กรอบราดหน้าทะเล	344
ผัดไทยกุ้งสด	519
ขนมจีนแกงเขียวหวาน	337

1) จงหาพิสัยและพิสัยระหว่างควอร์ไทล์ของปริมาณพลังงานของอาหารจานเดียวที่สุ่มตัวอย่างมา 11 รายการนี้
2) จงพิจารณาว่าระหว่างพิสัยและพิสัยระหว่างควอร์ไทล์ ค่าวัดการกระจายสัมบูรณ์ใดเหมาะสำหรับใช้อธิบายลักษณะการกระจายของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ

3. ครอบครัวหนึ่งประกอบด้วยพ่อ แม่ และลูกอีก 3 คน มีอายุ 45, 42, 20, 17 และ 16 ปี ตามลำดับ จงหาส่วนเบี่ยงเบนมาตรฐานและความแปรปรวนของอายุของสมาชิกในครอบครัวนี้ และจงหาว่าในอีก 5 ปีข้างหน้า ส่วนเบี่ยงเบนมาตรฐานของอายุของสมาชิกในครอบครัวนี้จะเป็นอย่างไร

4. จากรายงานของศูนย์ข้อมูลอุบัติเหตุ เพื่อเสริมสร้างวัฒนธรรมความปลอดภัยทางถนน พบว่าจำนวนผู้บาดเจ็บรวม (ราย) ตั้งแต่ พ.ศ. 2556 – 2558 ในแต่ละวันของช่วง 7 วันอันตรายของเทศกาลปีใหม่ แสดงได้ดังนี้

วันที่ 1	วันที่ 2	วันที่ 3	วันที่ 4	วันที่ 5	วันที่ 6	วันที่ 7
1,236	1,633	1,664	1,458	1,506	1,423	870

จงหาพิสัย พิสัยระหว่างควอร์ไทล์ และส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้

5. ค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 2 ห้องเรียน ซึ่งมีคะแนนเต็ม 100 คะแนน เป็นดังนี้

	ค่าเฉลี่ยเลขคณิต	ส่วนเบี่ยงเบนมาตรฐาน
ห้อง 1	73.2	4.8
ห้อง 2	52.4	3.6

จงเปรียบเทียบการกระจายของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนสองห้องนี้

6. อุณหภูมิสูงสุดและอุณหภูมิต่ำสุด (องศาเซลเซียส) ของจังหวัดขอนแก่น ตั้งแต่ พ.ศ. 2549 – 2558 แสดงได้ดังนี้

พ.ศ.	2549	2550	2551	2552	2553	2554	2555	2556	2557	2558
อุณหภูมิสูงสุด	39.3	41.1	38.5	39.6	41.2	39.3	39.0	41.8	40.5	41.0
อุณหภูมิต่ำสุด	12.0	12.6	11.9	10.2	13.5	11.6	15.0	11.6	10.2	11.6

จงหาสัมประสิทธิ์การแปรผันของอุณหภูมิสูงสุดและอุณหภูมิต่ำสุดของจังหวัดขอนแก่น ตั้งแต่ พ.ศ. 2549 – 2558 พร้อมทั้งเปรียบเทียบการกระจายของข้อมูลทั้งสองชุดนี้

3.3.3 ค่าวัดตำแหน่งที่ของข้อมูล

การวัดตำแหน่งที่ของข้อมูลเป็นการพิจารณาตำแหน่งที่ของข้อมูลตัวหนึ่งเมื่อเปรียบเทียบกับข้อมูลตัวอื่น ๆ ที่อยู่ในชุดข้อมูลเดียวกัน เช่น จากผลการแข่งขันคณิตศาสตร์โอลิมปิกระหว่างประเทศ (International Mathematical Olympiad: IMO) ครั้งที่ 55 พ.ศ. 2557 พบว่าประเทศไทยอยู่ในอันดับที่ 21 ถ้าไม่ได้เปรียบเทียบอันดับที่ของประเทศไทยกับประเทศที่เข้าร่วมการแข่งขันทั้งหมด จะไม่สามารถทราบได้ว่าศักยภาพทางด้านคณิตศาสตร์ของผู้แทนประเทศไทยเป็นอย่างไร เมื่อเปรียบเทียบกับประเทศอื่น ๆ ที่เข้าร่วมการแข่งขัน แต่ถ้ามีการเปรียบเทียบอันดับที่ของประเทศไทยกับประเทศที่เข้าร่วมการแข่งขันทั้งหมด 101 ประเทศ จะเห็นว่าผู้แทนประเทศไทยทำผลงานได้ดีมากจนติดอันดับต้น ๆ ของโลก

ค่าวัดตำแหน่งที่ของข้อมูลที่นิยมใช้กันมาก คือ ควอร์ไทล์และเปอร์เซ็นไทล์

ควอร์ไทล์ (Quartile)

นักเรียนได้ศึกษาการหาตำแหน่งของควอร์ไทล์มาแล้วในหัวข้อ 3.2 เรื่องแผนภาพกล่อง ซึ่งควอร์ไทล์มีทั้งหมดสามค่า ได้แก่ ควอร์ไทล์ที่ 1 (Q_1) ควอร์ไทล์ที่ 2 (Q_2) และควอร์ไทล์ที่ 3 (Q_3) โดยควอร์ไทล์จะแบ่งข้อมูลที่เรียงจากน้อยไปมากออกเป็น 4 ส่วน เท่า ๆ กัน ดังรูปที่ 7 จะเห็นว่า ควอร์ไทล์ที่ i (Q_i) เมื่อ $i \in \{1, 2, 3\}$ เป็นค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณ i ส่วน และมีจำนวนข้อมูลที่มีค่ามากกว่าค่านี้อยู่ประมาณ $4-i$ ส่วน



รูปที่ 7

ให้ n แทนจำนวนข้อมูลทั้งหมด และ $i \in \{1, 2, 3\}$ การหาควอร์ไทล์ที่ i (Q_i) ทำได้โดยเรียงลำดับข้อมูล n ตัว จากน้อยไปมาก จากนั้นจะได้ว่า Q_i อยู่ในตำแหน่งที่ $\frac{i(n+1)}{4}$

ข้อสังเกต เนื่องจากมัธยฐาน คือ Q_2 ดังนั้น มัธยฐานจึงเป็นค่ากลางที่สามารถใช้ในการวัดตำแหน่งที่ของข้อมูลได้

ตัวอย่างที่ 23

ข้อมูลปริมาณการส่งออกข้าวไทยโดยประมาณ (พันตันข้าวสาร) ใน พ.ศ. 2560 จำแนกตามชนิดของข้าว จากศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงพาณิชย์ โดยความร่วมมือจากกรมศุลกากร แสดงได้ดังตาราง

ชนิดของข้าว	ปริมาณการส่งออกโดยประมาณ (พันตันข้าวสาร)
ต้นข้าวขาว	4,662
ปลายข้าวขาว	408
ต้นข้าวหอมมะลิ	1,630
ปลายข้าวหอมมะลิ	669
ข้าวนึ่ง	3,370
ข้าวเหนียว	214
ปลายข้าวเหนียว	303
ข้าวหอมไทย	213

จงหา

1) ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด ใน พ.ศ. 2560
2) ชนิดของข้าวที่มีปริมาณการส่งออกน้อยกว่าควอร์ไทล์ที่ 1
3) ชนิดของข้าวที่มีปริมาณการส่งออกมากกว่าควอร์ไทล์ที่ 3

วิธีทำ เรียงปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด จากน้อยไปมาก ได้ดังนี้

213 214 303 408 669 1,630 3,370 4,662

1) เนื่องจาก Q_1 อยู่ในตำแหน่งที่ $\frac{1(8+1)}{4} = 2.25$
ดังนั้น Q_1 อยู่ระหว่างข้อมูลในตำแหน่งที่ 2 และ 3 ซึ่งมีค่าอยู่ระหว่าง 214 และ 303
ในการหา Q_1 จะใช้การเทียบบัญญัติไตรยางศ์ ดังนี้
เนื่องจากข้อมูลในตำแหน่งที่ 2 และ 3 มีตำแหน่งต่างกัน $3-2=1$ มีค่าต่างกัน $303-214=89$
จะได้ว่าตำแหน่งต่างกัน $2.25-2=0.25$ มีค่าต่างกัน $\frac{0.25\times 89}{1} = 22.25$
ดังนั้น $Q_1 = 214+22.25 = 236.25$
เนื่องจาก Q_2 อยู่ในตำแหน่งที่ $\frac{2(8+1)}{4} = 4.5$
ดังนั้น Q_2 คือ ค่าเฉลี่ยเลขคณิตของข้อมูลในตำแหน่งที่ 4 และ 5 ซึ่งคือ $\frac{408+669}{2} = 538.5$
เนื่องจาก Q_3 อยู่ในตำแหน่งที่ $\frac{3(8+1)}{4} = 6.75$
ดังนั้น Q_3 อยู่ระหว่างข้อมูลในตำแหน่งที่ 6 และ 7 ซึ่งมีค่าอยู่ระหว่าง 1,630 และ 3,370
ในการหา Q_3 จะใช้การเทียบบัญญัติไตรยางศ์ ดังนี้
เนื่องจากข้อมูลในตำแหน่งที่ 6 และ 7 มีตำแหน่งต่างกัน $7-6=1$ มีค่าต่างกัน $3,370-1,630=1,740$
จะได้ว่าตำแหน่งต่างกัน $6.75-6=0.75$ มีค่าต่างกัน $\frac{0.75\times 1,740}{1} = 1,305$
ดังนั้น $Q_3 = 1,630+1,305 = 2,935$
จะได้ว่าควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของปริมาณการส่งออกโดยประมาณของข้าว 8 ชนิด ใน พ.ศ. 2560 คือ 236.25, 538.5 และ 2,935 พันตันข้าวสาร ตามลำดับ

2) ชนิดของข้าวที่มีปริมาณการส่งออกน้อยกว่าควอร์ไทล์ที่ 1 ซึ่งเท่ากับ 236.25 พันตันข้าวสาร ได้แก่ ข้าวหอมไทย และข้าวเหนียว

3) ชนิดของข้าวที่มีปริมาณการส่งออกมากกว่าควอร์ไทล์ที่ 3 ซึ่งเท่ากับ 2,935 พันตันข้าวสาร ได้แก่ ข้าวนึ่ง และต้นข้าวขาว ■

เปอร์เซ็นไทล์ (Percentile)

ในทำนองเดียวกันกับควอร์ไทล์ การวัดตำแหน่งโดยเปอร์เซ็นไทล์เป็นการแบ่งข้อมูลที่เรียงจากน้อยไปมาก แต่เปอร์เซ็นไทล์แบ่งข้อมูลทั้งหมดออกเป็น 100 ส่วน เท่า ๆ กัน จึงประกอบด้วยเปอร์เซ็นไทล์ที่ 1, 2, 3, … , 99 โดยเปอร์เซ็นไทล์ที่ i เมื่อ $i \in \{1, 2, 3, \dots, 99\}$ แทนด้วยสัญลักษณ์ P_i หมายความว่าเมื่อแบ่งข้อมูลที่เรียงจากน้อยไปมากออกเป็น 100 ส่วน เท่า ๆ กัน เปอร์เซ็นไทล์ที่ i (P_i) เมื่อ $i \in \{1, 2, 3, \dots, 99\}$ จะเป็นค่าที่มีจำนวนข้อมูลที่มีค่าน้อยกว่าค่านี้อยู่ประมาณ i ส่วน หรือร้อยละ i ของข้อมูลทั้งหมด และมีจำนวนข้อมูลที่มีค่ามากกว่าค่านี้อยู่ประมาณ $100 - i$ ส่วน หรือร้อยละ $100 - i$ ของข้อมูลทั้งหมด

ให้ n แทนจำนวนข้อมูลทั้งหมด และ $i \in \{1, 2, 3, \dots, 99\}$ การหาเปอร์เซ็นไทล์ที่ i (P_i) ทำได้โดยเรียงลำดับข้อมูล n ตัว จากน้อยไปมาก จากนั้นจะได้ว่า P_i อยู่ในตำแหน่งที่ $\frac{i(n+1)}{100}$

หมายเหตุ พิจารณาตำแหน่งที่ของ Q_1, Q_2 และ Q_3 กับตำแหน่งที่ของ P_{25}, P_{50} และ P_{75} ตามลำดับ จะได้ว่า

- ตำแหน่งที่ของ Q_1 คือ $\frac{1(n+1)}{4}$ ซึ่งเท่ากับ $\frac{25(n+1)}{100}$ จึงเป็นตำแหน่งที่เดียวกันกับ P_{25}
- ตำแหน่งที่ของ Q_2 คือ $\frac{2(n+1)}{4}$ ซึ่งเท่ากับ $\frac{50(n+1)}{100}$ จึงเป็นตำแหน่งที่เดียวกันกับ P_{50}
- ตำแหน่งที่ของ Q_3 คือ $\frac{3(n+1)}{4}$ ซึ่งเท่ากับ $\frac{75(n+1)}{100}$ จึงเป็นตำแหน่งที่เดียวกันกับ P_{75}

นั่นคือ $Q_1 = P_{25}, Q_2 = P_{50}$ และ $Q_3 = P_{75}$

ตัวอย่างที่ 24

คะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1 ในโรงเรียนแห่งหนึ่ง จำนวนทั้งหมด 40 คน ซึ่งมีคะแนนเต็ม 100 คะแนน แสดงได้ดังนี้

96	78	80	76	84	77	74	85	65	69
82	53	45	67	58	54	56	62	56	54
43	48	49	50	60	65	54	51	55	60
65	66	75	98	97	63	92	94	76	78

จงหา

1) เปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 80 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้
2) คะแนนที่มีนักเรียนประมาณร้อยละ 25 ของห้องได้คะแนนต่ำกว่า
3) คะแนนที่มีนักเรียนประมาณหนึ่งในห้าของห้องได้คะแนนสูงกว่า

วิธีทำ เรียงคะแนนสอบวิชาภาษาอังกฤษของนักเรียน 40 คน จากน้อยไปมาก ได้ดังนี้

43	45	48	49	50	51	53	54	54	54
55	56	56	58	60	60	62	63	65	65
65	66	67	69	74	75	76	76	77	78
78	80	82	84	85	92	94	96	97	98

1) เนื่องจาก P_{25} อยู่ในตำแหน่งที่ $\frac{25(40+1)}{100} = 10.25$

ดังนั้น P_{25} อยู่ระหว่างข้อมูลในตำแหน่งที่ 10 และ 11 ซึ่งมีค่าอยู่ระหว่าง 54 และ 55

ในการหา P_{25} จะใช้การเทียบบัญญัติไตรยางศ์ ดังนี้

เนื่องจากข้อมูลในตำแหน่งที่ 10 และ 11 มีตำแหน่งต่างกัน $11-10=1$ มีค่าต่างกัน $55-54=1$

จะได้ว่าตำแหน่งต่างกัน $10.25-10=0.25$ มีค่าต่างกัน $\frac{0.25\times 1}{1} = 0.25$

ดังนั้น $P_{25} = 54 + 0.25 = 54.25$

เนื่องจาก P_{50} อยู่ในตำแหน่งที่ $\frac{50(40+1)}{100} = 20.5$

ดังนั้น P_{50} อยู่ระหว่างข้อมูลในตำแหน่งที่ 20 และ 21 ซึ่งมีค่าอยู่ระหว่าง 65 และ 65 นั่นคือ $P_{50} = 65$

เนื่องจาก P_{75} อยู่ในตำแหน่งที่ $\frac{75(40+1)}{100} = 30.75$

ดังนั้น P_{75} อยู่ระหว่างข้อมูลในตำแหน่งที่ 30 และ 31 ซึ่งมีค่าอยู่ระหว่าง 78 และ 78 นั่นคือ $P_{75} = 78$

เนื่องจาก P_{80} อยู่ในตำแหน่งที่ $\frac{80(40+1)}{100} = 32.8$

ดังนั้น P_{80} อยู่ระหว่างข้อมูลในตำแหน่งที่ 32 และ 33 ซึ่งมีค่าอยู่ระหว่าง 80 และ 82 ในการหา P_{80} จะใช้การเทียบบัญญัติไตรยางศ์ ดังนี้

เนื่องจากข้อมูลในตำแหน่งที่ 32 และ 33 มีตำแหน่งต่างกัน $33 - 32 = 1$ มีค่าต่างกัน $82 - 80 = 2$

จะได้ว่าตำแหน่งต่างกัน $32.8 - 32 = 0.8$ มีค่าต่างกัน $\frac{0.8 \times 2}{1} = 1.6$

ดังนั้น $P_{80} = 80 + 1.6 = 81.6$

จะได้ว่าเปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 80 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้ คือ 54.25, 65, 78 และ 81.6 คะแนน ตามลำดับ

2) คะแนนที่มีนักเรียนประมาณร้อยละ 25 ของห้องได้คะแนนต่ำกว่า คือ P_{25} ซึ่งเท่ากับ 54.25 คะแนน

3) คะแนนที่มีนักเรียนประมาณหนึ่งในห้าของห้องได้คะแนนสูงกว่า คือ P_{80} ซึ่งเท่ากับ 81.6 คะแนน ■

ควอร์ไทล์และเปอร์เซ็นไทล์สามารถใช้ในการวิเคราะห์ข้อมูลได้ ถึงแม้ว่าข้อมูลชุดนั้นจะมีค่านอกเกณฑ์ก็ไม่มีผลต่อควอร์ไทล์และเปอร์เซ็นไทล์ เนื่องจากการหาควอร์ไทล์และเปอร์เซ็นไทล์จะพิจารณาเพียงตำแหน่งที่ของข้อมูลเท่านั้น

แบบฝึกหัด 3.3.3

1. โรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่จำนวน 15 แห่ง มีกำลังผลิต (เมกะวัตต์) ในเดือนมกราคม พ.ศ. 2562 ดังนี้

เขื่อน	กำลังผลิต (เมกะวัตต์)
ภูมิพล	779.20
สิริกิติ์	500.00
อุบลรัตน์	25.20
สิรินธร	36.00
จุฬาภรณ์	40.00
ศรีนครินทร์	720.00
วชิราลงกรณ	300.00
ท่าทุ่งนา	39.00
แก่งกระจาน	19.00
บางลาง	84.00
รัชชประภา	240.00
ปากมูล	136.00
เจ้าพระยา	12.00
แควน้อยบำรุงแดน	30.00
แม่กลอง	12.00

จงหา

1) ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) กำลังผลิตที่มีโรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่จำนวนประมาณครึ่งหนึ่งมีกำลังผลิตได้น้อยกว่า
3) โรงไฟฟ้าพลังน้ำเขื่อนขนาดใหญ่ที่มีกำลังผลิตมากกว่าควอร์ไทล์ที่ 3

2. โรงพยาบาลแห่งหนึ่งบันทึกจำนวนทารกแรกเกิดตั้งแต่เดือนมกราคม – ธันวาคม พ.ศ. 2562 ได้ข้อมูลดังนี้

เดือน	จำนวนทารกแรกเกิด (คน)
มกราคม	305
กุมภาพันธ์	289
มีนาคม	313
เมษายน	342
พฤษภาคม	311
มิถุนายน	324
กรกฎาคม	345
สิงหาคม	341
กันยายน	353
ตุลาคม	329
พฤศจิกายน	304
ธันวาคม	324

จงหา

1) ควอร์ไทล์ที่ 1 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) เดือนที่มีจำนวนทารกแรกเกิดน้อยกว่าควอร์ไทล์ที่ 1
3) เดือนที่มีจำนวนทารกแรกเกิดมากกว่าควอร์ไทล์ที่ 3

3. ข้อมูลระยะเวลาตั้งท้องเฉลี่ย (วัน) และอายุขัยเฉลี่ย (ปี) ของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิด แสดงได้ดังนี้

สัตว์เลี้ยงลูกด้วยน้ำนม	ระยะเวลาตั้งท้องเฉลี่ย (วัน)	อายุขัยเฉลี่ย (ปี)
สิงโต	100	15
ลิง	166	15
ม้าลาย	365	15
เสือ	105	16
กวาง	201	8
ฮิปโปโปเตมัส	238	41
ช้าง	660	35
ยีราฟ	425	10
อูฐ	406	12
ม้า	330	20

1) จงหาเปอร์เซ็นไทล์ที่ 20 และเปอร์เซ็นไทล์ที่ 80 ของระยะเวลาตั้งท้องเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้
2) จงหาเปอร์เซ็นไทล์ที่ 20 และเปอร์เซ็นไทล์ที่ 80 ของอายุขัยเฉลี่ยของสัตว์เลี้ยงลูกด้วยน้ำนม 10 ชนิดนี้
3) ถ้าสวนสัตว์แห่งหนึ่งต้องการเพิ่มจำนวนสัตว์ โดยจะเลือกจากสัตว์ที่มีระยะเวลาตั้งท้องเฉลี่ยมากกว่าเปอร์เซ็นไทล์ที่ 80 หรือสัตว์ที่มีอายุขัยเฉลี่ยน้อยกว่าเปอร์เซ็นไทล์ที่ 20 สวนสัตว์แห่งนี้จะเลือกเพิ่มจำนวนสัตว์ชนิดใดได้บ้าง

4. จากตัวอย่างที่ 24 ถ้าคะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/2 ในโรงเรียนเดียวกัน จำนวนทั้งหมด 30 คน ซึ่งมีคะแนนเต็ม 100 คะแนน แสดงได้ดังนี้

88	70	61	43	31	56	64	82	73	67
55	73	57	37	78	77	59	35	27	86
61	49	54	60	74	49	78	68	70	78

ข้อสรุปที่ว่า "เปอร์เซ็นไทล์ที่ 90 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/2 มากกว่าเปอร์เซ็นไทล์ที่ 80 ของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 6/1" เป็นจริงหรือไม่

5. จากการสำรวจของสำนักงานคณะกรรมการข้าราชการพลเรือน (ก.พ.) และสำนักงานสถิติแห่งชาติใน พ.ศ. 2556 พบว่า เงินเดือนของพนักงานใหม่แรกบรรจุ จำแนกตามวุฒิการศึกษาและตำแหน่ง แสดงได้ดังตารางต่อไปนี้

ตำแหน่ง	เงินเดือน (บาท)	
	ปริญญาตรี	ปริญญาโท/เอก
เจ้าพนักงานธุรการ เจ้าหน้าที่ธุรการ และพนักงานพิมพ์ดีด	12,166	-
เจ้าพนักงานการเงินและบัญชี เจ้าหน้าที่บัญชี และเจ้าหน้าที่การเงิน	13,184	19,940
นักวิเคราะห์นโยบายและแผน และนักวางแผน	14,308	22,643
นักทรัพยากรบุคคล เจ้าหน้าที่ฝึกอบรม เจ้าหน้าที่ทรัพยากรบุคคล และเจ้าหน้าที่วิเทศสัมพันธ์	13,219	20,957
นักประชาสัมพันธ์ และเจ้าหน้าที่ประชาสัมพันธ์	12,760	18,028

ตำแหน่ง	เงินเดือน (บาท)	
	ปริญญาตรี	ปริญญาโท/เอก
นักวิชาการคอมพิวเตอร์ และนักเขียนโปรแกรม	15,263	21,342
นิติกร และนักกฎหมาย	16,000	23,823
เศรษฐกร และนักเศรษฐศาสตร์	15,043	21,050
วิศวกรเหมืองแร่ วิศวกรไฟฟ้า และเครื่องกล	16,986	24,163
นักสำรวจ (ปิโตรเลียม) และวิศวกรปิโตรเลียม	17,829	26,686
สถาปนิก	18,266	28,677
นักทรัพยากรธรณี	18,039	26,555
นักวิชาการขนส่ง และนักโลจิสติกส์	15,930	23,125
นักวิทยาศาสตร์ และนักเคมี	16,138	23,454
แพทย์	-	63,082
ทันตแพทย์	-	56,807
พยาบาล	16,487	39,526
เภสัชกร	20,003	33,050
นักเทคนิคการแพทย์	16,063	32,505
นักรังสีการแพทย์	16,267	37,871
นักโภชนาการ	14,973	30,879

1) เงินเดือนของพนักงานใหม่แรกบรรจุที่มีวุฒิปริญญาตรีหรือวุฒิปริญญาโท/เอก มีการกระจายมากกว่ากัน
2) สำหรับพนักงานใหม่แรกบรรจุที่มีวุฒิปริญญาโท/เอก มีตำแหน่งใดบ้างที่ได้เงินเดือนน้อยกว่าเปอร์เซ็นไทล์ที่ 30

กิจกรรม : ความล่าช้าของเที่ยวบิน

สายการบินแห่งหนึ่งต้องการปรับปรุงการให้บริการของเที่ยวบินจากขอนแก่นไปกรุงเทพฯ จึงได้เก็บข้อมูลระยะเวลาที่เที่ยวบินล่าช้า (นาที) ในเดือนเมษายน พ.ศ. 2560 ถึงเดือนกุมภาพันธ์ พ.ศ. 2561 จำนวนทั้งหมด 170 เที่ยวบิน แสดงได้ดังนี้

เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า	เที่ยวบินที่	ระยะเวลาที่ล่าช้า
1	26	23	25	45	10	67	14	89	165	111	4	133	7	155	12
2	20	24	26	46	16	68	12	90	15	112	3	134	8	156	2
3	27	25	15	47	3	69	29	91	24	113	9	135	9	157	11
4	26	26	8	48	14	70	5	92	22	114	5	136	10	158	8
5	24	27	21	49	23	71	8	93	13	115	21	137	12	159	11
6	22	28	7	50	94	72	9	94	27	116	10	138	6	160	8
7	9	29	19	51	5	73	6	95	13	117	2	139	14	161	8
8	27	30	13	52	33	74	15	96	15	118	3	140	4	162	6
9	14	31	16	53	22	75	19	97	7	119	1	141	5	163	9
10	18	32	14	54	12	76	15	98	22	120	7	142	7	164	10
11	19	33	19	55	6	77	11	99	13	121	11	143	15	165	137
12	16	34	26	56	47	78	7	100	8	122	5	144	6	166	9
13	14	35	16	57	29	79	8	101	10	123	27	145	7	167	3
14	21	36	28	58	14	80	24	102	20	124	4	146	13	168	10
15	14	37	8	59	11	81	27	103	3	125	9	147	4	169	7
16	18	38	10	60	30	82	19	104	2	126	41	148	9	170	9
17	33	39	16	61	26	83	37	105	1	127	2	149	4		
18	17	40	13	62	9	84	23	106	1	128	6	150	7		
19	33	41	12	63	16	85	24	107	4	129	10	151	15		
20	18	42	12	64	13	86	27	108	3	130	3	152	8		
21	35	43	19	65	18	87	23	109	2	131	12	153	8		
22	25	44	16	66	19	88	13	110	2	132	9	154	6		

นักเรียนสามารถคัดลอกชุดข้อมูลนี้ได้ที่ ipst.me/10680

ขั้นตอนการปฏิบัติ

1. เขียนตารางความถี่ของข้อมูลชุดนี้
2. นำเสนอข้อมูลชุดนี้ด้วยรูปแบบการนำเสนอที่คิดว่าเหมาะสม พร้อมทั้งให้เหตุผลประกอบ
3. หาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลชุดนี้ และพิจารณาว่าควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ
4. หาพิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนของข้อมูลชุดนี้ พร้อมทั้งอธิบายความหมายของค่าที่คำนวณได้
5. หาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
6. หาค่านอกเกณฑ์ของข้อมูลชุดนี้
7. ถ้าตัดค่านอกเกณฑ์ออกจากข้อมูลชุดนี้ จงหาค่าเฉลี่ยเลขคณิตของข้อมูลที่เหลือ
8. แทนค่านอกเกณฑ์ของข้อมูลชุดนี้ด้วยค่าเฉลี่ยเลขคณิตที่หาได้ในข้อ 7 แล้วพิจารณาว่า
   8.1 ค่ากลางใดได้รับผลกระทบจากการแทนข้อมูลดังกล่าวมากที่สุด เพราะเหตุใด
   8.2 ค่าวัดการกระจายใดได้รับผลกระทบจากการแทนข้อมูลดังกล่าวมากที่สุด เพราะเหตุใด
   8.3 ควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ได้รับผลกระทบจากการแทนข้อมูลดังกล่าวหรือไม่ เพราะเหตุใด

แบบฝึกหัดท้ายบท

1 ยอดชำระเงิน (บาท) ของลูกค้าที่มาซื้อของในซูเปอร์มาร์เก็ตของห้างสรรพสินค้าแห่งหนึ่ง จำนวน 50 คน แสดงได้ดังนี้

217	352	592	156	421	261	508	323	236	430
301	291	439	524	287	460	341	137	369	262
402	316	89	372	201	423	417	468	248	672
209	409	347	446	387	410	336	360	581	112
456	221	546	256	391	182	454	50	446	352

1) จงเขียนตารางความถี่พร้อมทั้งแสดงความถี่สะสม ความถี่สัมพัทธ์ และความถี่สะสมสัมพัทธ์ของข้อมูลชุดนี้ โดยกำหนดอันตรภาคชั้นเป็น ต่ำกว่า 100, 100 – 199, 200 – 299, 300 – 399, 400 – 499, 500 – 599 และ 600 – 699
2) ลูกค้ามียอดชำระเงินอยู่ในอันตรภาคชั้นใดมากที่สุด
3) ลูกค้าที่มียอดชำระเงินต่ำกว่า 100 บาท มีจำนวนมากหรือน้อยกว่าลูกค้าที่มียอดชำระเงินตั้งแต่ 600 – 699 บาท
4) ลูกค้าที่มียอดชำระเงินตั้งแต่ 400 บาทขึ้นไป คิดเป็นร้อยละเท่าใดของจำนวนลูกค้า 50 คนนี้
5) ลูกค้าที่มียอดชำระเงินตั้งแต่ 200 บาท แต่น้อยกว่า 500 บาท คิดเป็นร้อยละเท่าใดของจำนวนลูกค้า 50 คนนี้

2 จากการสำรวจปริมาณฟลูออไรด์ (มิลลิกรัมต่อลิตร) ในน้ำดื่มบรรจุขวด จากตัวอย่างน้ำดื่มบรรจุขวดที่จำหน่ายในกรุงเทพมหานครและปริมณฑล จำนวน 30 ยี่ห้อ ได้ข้อมูลดังนี้

ยี่ห้อ	ปริมาณฟลูออไรด์ (มิลลิกรัมต่อลิตร)	ยี่ห้อ	ปริมาณฟลูออไรด์ (มิลลิกรัมต่อลิตร)
1	0.63	16	4.24
2	0.74	17	1.30
3	0.85	18	0.07
4	0.14	19	0.76
5	0.06	20	0.00
6	0.01	21	1.08
7	1.60	22	0.13
8	2.81	23	0.03
9	0.36	24	0.08
10	0.75	25	0.49
11	1.00	26	0.31
12	2.52	27	0.42
13	0.50	28	0.77
14	0.00	29	0.31
15	0.21	30	1.05

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดอันตรภาคชั้นเป็น $0 \le x < 0.5$, $0.5 \le x < 1$, $1 \le x < 1.5$, $1.5 \le x < 2$ และ $x \ge 2$ เมื่อ x แทนปริมาณฟลูออไรด์ (มิลลิกรัมต่อลิตร) พร้อมทั้งสรุปผลที่ได้
2) จำนวนยี่ห้อน้ำดื่มบรรจุขวดที่มีปริมาณฟลูออไรด์ตั้งแต่ 2 มิลลิกรัมต่อลิตรขึ้นไป คิดเป็นร้อยละเท่าใดของจำนวนยี่ห้อน้ำดื่มบรรจุขวดที่สำรวจทั้งหมด

เสริมสมอง : ฟลูออไรด์

ฟลูออไรด์ ไอออน (fluoride ion) หรือบางครั้งเรียกสั้น ๆ ว่า ฟลูออไรด์ (fluoride) เป็นไอออนของธาตุฟลูออรีน ซึ่งในธรรมชาติอยู่ในรูปของสารประกอบ พบได้ทั่วไปทั้งในดิน น้ำ อากาศ หินแร่ และพืชบางชนิด การได้รับฟลูออไรด์ในปริมาณที่เหมาะสมจะช่วยป้องกันโรคฟันผุ แต่การได้รับฟลูออไรด์มากเกินไปตั้งแต่เด็กแรกเกิดจนถึง 8 ปี จะทำให้เกิดฟันตกกระ ซึ่งเป็นภาวะที่ฟันมีสีขาวขุ่นเหมือนมีลาย ถ้าเป็นรุนแรงฟันจะมีสีน้ำตาลหรือดำ และเปราะง่าย สาเหตุส่วนใหญ่มาจากการบริโภคน้ำที่มีปริมาณฟลูออไรด์เกินมาตรฐาน โดยกรมอนามัยได้กำหนดมาตรฐานปริมาณฟลูออไรด์ในน้ำบริโภคว่าต้องไม่มากกว่า 0.7 มิลลิกรัมต่อลิตร



3 ปริมาณน้ำฝนรายเดือน (มิลลิเมตร) ของจังหวัดระยองตั้งแต่ พ.ศ. 2554 – 2556 แสดงได้ดังนี้

	ปริมาณน้ำฝน (มิลลิเมตร)		
	พ.ศ. 2554	พ.ศ. 2555	พ.ศ. 2556
มกราคม	0.0	82.8	116.4
กุมภาพันธ์	65.2	80.2	5.0
มีนาคม	150.6	60.3	36.3
เมษายน	102.7	8.9	96.9
พฤษภาคม	33.7	184.8	130.0
มิถุนายน	278.5	66.7	338.1

	ปริมาณน้ำฝน (มิลลิเมตร)		
	พ.ศ. 2554	พ.ศ. 2555	พ.ศ. 2556
กรกฎาคม	143.6	258.4	212.7
สิงหาคม	288.0	64.7	166.4
กันยายน	373.8	504.7	40.4
ตุลาคม	194.8	202.0	295.7
พฤศจิกายน	18.9	59.3	69.1
ธันวาคม	0.6	0.0	11.4

จงเขียนตารางความถี่ของข้อมูลนี้ โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 5 ชั้น ค่าเริ่มต้นเท่ากับ 0 มิลลิเมตร และค่าสุดท้ายเท่ากับ 510 มิลลิเมตร พร้อมทั้งสรุปผลที่ได้

4 คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง แสดงด้วยตารางความถี่ได้ดังนี้

คะแนน	จำนวนนักเรียน (คน)
30 – 39	2
40 – 49	0
50 – 59	6
60 – 69	6
70 – 79	10
80 – 89	13
90 – 99	8
รวม	**45**

ถ้ากำหนดเกณฑ์ในการแบ่งระดับคะแนน ดังนี้

เกรด	คะแนน
4	90 – 100
3	80 – 89
2	70 – 79
1	50 – 69
0	ต่ำกว่า 50

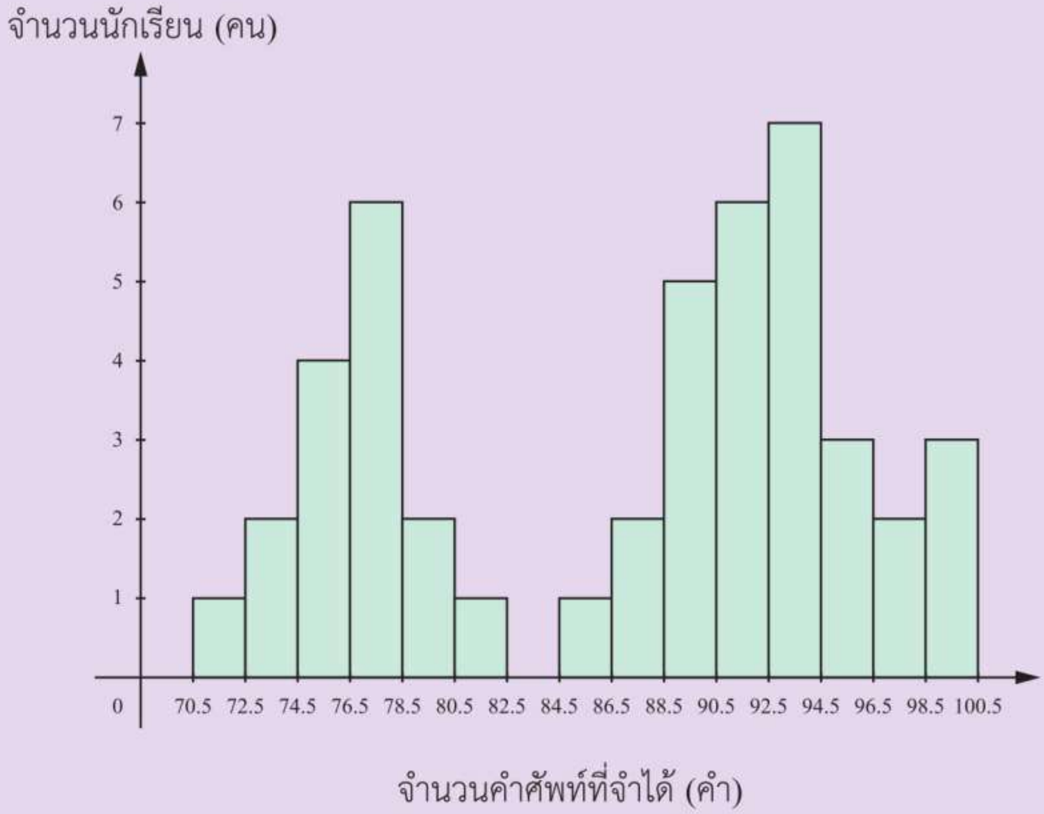
1) จงเขียนตารางความถี่ของข้อมูลชุดนี้ โดยกำหนดอันตรภาคชั้นเป็นระดับคะแนนที่กำหนดในเกณฑ์ข้างต้น
2) นักเรียนห้องนี้ได้เกรดใดมากที่สุด
3) นักเรียนที่ได้ตั้งแต่เกรด 3 ขึ้นไป คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด

5 ร้านค้าออนไลน์แห่งหนึ่งมีนโยบายการรับประกันความพึงพอใจ โดยลูกค้าสามารถส่งคืนสินค้าที่ยังไม่ได้ใช้งานภายใน 7 วัน นับจากวันที่ได้รับสินค้า เพื่อรับเงินคืนเต็มจำนวน ถ้าข้อมูลจำนวนสินค้า (ชิ้น) ที่ลูกค้าส่งคืนในแต่ละวัน ในระยะเวลา 30 วัน แสดงได้ดังนี้

4	9	5	6	5	8	22	3	8	4
8	2	7	19	12	3	5	16	3	6
5	15	9	14	4	13	7	2	9	7

1) จงเขียนตารางความถี่ของข้อมูลชุดนี้พร้อมทั้งแสดงขอบล่างและขอบบนของแต่ละชั้น โดยกำหนดให้จำนวนอันตรภาคชั้นเท่ากับ 5 ชั้น ค่าเริ่มต้นเท่ากับ 0 ชิ้น และค่าสุดท้ายเท่ากับ 25 ชิ้น
2) จงเขียนฮิสโทแกรมของข้อมูลชุดนี้
3) จงเขียนแผนภาพจุดของข้อมูลชุดนี้
4) จงเขียนแผนภาพกล่องของข้อมูลชุดนี้

6. จากการทดสอบเทคนิคการจำคำศัพท์ใหม่กับนักเรียนที่เรียนภาษาจีนกลุ่มหนึ่งจำนวน 45 คน โดยให้นักเรียนใช้เทคนิคดังกล่าวในการจดจำคำศัพท์ใหม่จำนวน 100 คำ แล้วบันทึกจำนวนคำศัพท์ที่จำได้ในวันรุ่งขึ้น ผลการทดสอบแสดงด้วยฮิสโทแกรมได้ดังนี้



1) นักเรียนที่จำคำศัพท์ได้มากกว่า 92 คำ คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด
2) นักเรียนที่จำคำศัพท์ได้มากกว่า 70 คำ แต่ไม่เกิน 80 คำ คิดเป็นร้อยละเท่าใดของจำนวนนักเรียนทั้งหมด
3) จำนวนคำศัพท์ที่นักเรียนจำนวนมากที่สุดจำได้อยู่ในช่วงใด

7 ผลการทดสอบการอ่านภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 1/1 และ 1/2 ซึ่งมีคะแนนเต็ม 25 คะแนน แสดงด้วยตารางความถี่ได้ดังนี้

คะแนน	จำนวนนักเรียน (คน)	
	ชั้นประถมศึกษาปีที่ 1/1	ชั้นประถมศึกษาปีที่ 1/2
20 – 22	1	0
17 – 19	8	11
14 – 16	12	28
11 – 13	15	11
8 – 10	11	0
5 – 7	3	0

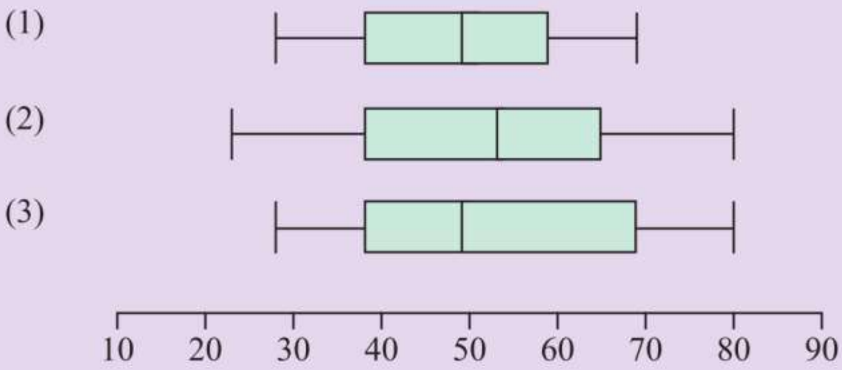
1) จงเขียนฮิสโทแกรมแสดงผลการทดสอบการอ่านภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 1/1
2) จงเขียนฮิสโทแกรมแสดงผลการทดสอบการอ่านภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 1/2
3) เมื่อพิจารณาฮิสโทแกรมของนักเรียนชั้นประถมศึกษาปีที่ 1/1 และ 1/2 คะแนนของนักเรียนห้องใดมีการกระจายมากกว่ากัน เพราะเหตุใด

8 สายการบินแห่งหนึ่งได้สำรวจความพึงพอใจของผู้โดยสารต่อการให้บริการของพนักงานต้อนรับบนเครื่องบิน ซึ่งมีคะแนนเต็ม 10 คะแนน โดยสำรวจจากผู้โดยสารที่สุ่มมาจำนวน 20 คน ได้ผลสำรวจโดยเรียงคะแนนจากน้อยไปมากดังนี้

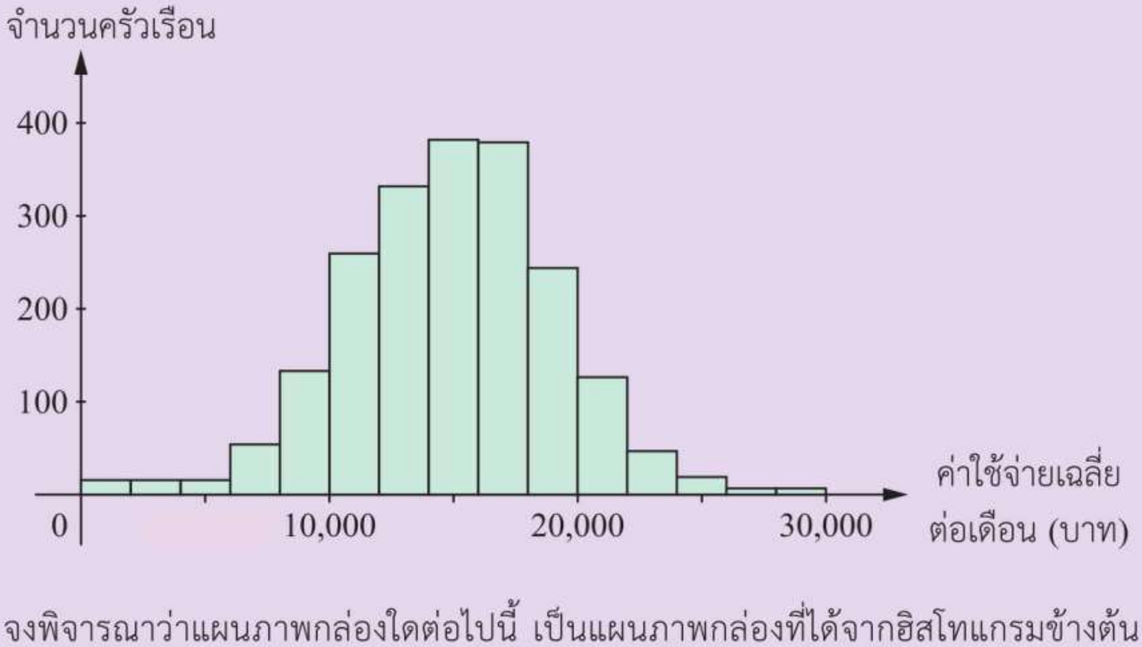
1 3 5 5 7 8 8 8 8 8 8 9 9 9 9 9 10 10 10 10

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) ข้อมูลชุดนี้มีค่านอกเกณฑ์หรือไม่ ถ้ามีคือค่าใด
3) จงเขียนแผนภาพกล่องเพื่อนำเสนอข้อมูลชุดนี้ พร้อมทั้งอธิบายลักษณะการกระจายของข้อมูลชุดนี้

9 ผลการสอบวิชาภาษาอังกฤษของนักเรียนกลุ่มหนึ่ง ซึ่งมีการสอบจำนวน 2 ครั้ง แต่ละครั้งมีคะแนนเต็ม 100 คะแนน แสดงด้วยแผนภาพกล่องได้ดังนี้ โดยแผนภาพกล่อง (1) แสดงคะแนนสอบครั้งที่ 1 และแผนภาพกล่อง (2) แสดงคะแนนสอบครั้งที่ 2 จงพิจารณาว่าเป็นไปได้หรือไม่ที่แผนภาพกล่อง (3) จะแสดงคะแนนเฉลี่ยจากการสอบทั้งสองครั้งของนักเรียนแต่ละคนในกลุ่มนี้

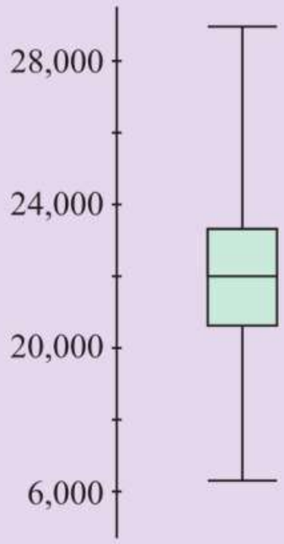


10 ผลการสำรวจค่าใช้จ่ายเฉลี่ยต่อเดือน (บาท) ของครัวเรือนในกรุงเทพมหานครที่เลือกมาเป็นตัวอย่าง แสดงด้วยฮิสโทแกรมได้ดังนี้

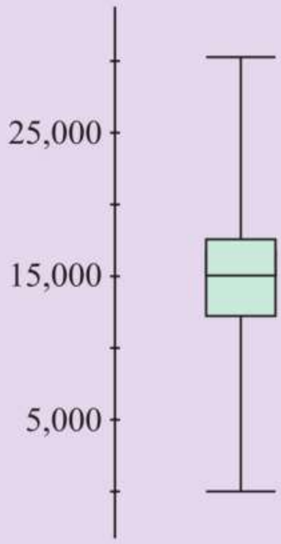


จงพิจารณาว่าแผนภาพกล่องใดต่อไปนี้ เป็นแผนภาพกล่องที่ได้จากฮิสโทแกรมข้างต้น พร้อมทั้งให้เหตุผลประกอบ

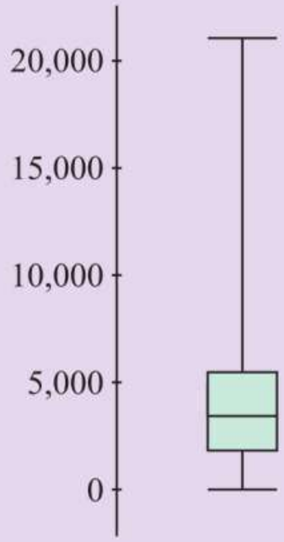
ก.


28,000
24,000
20,000
6,000

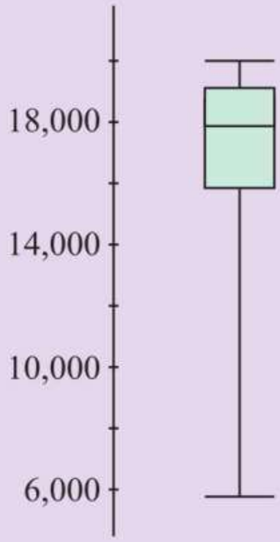
ข.


25,000
15,000
5,000

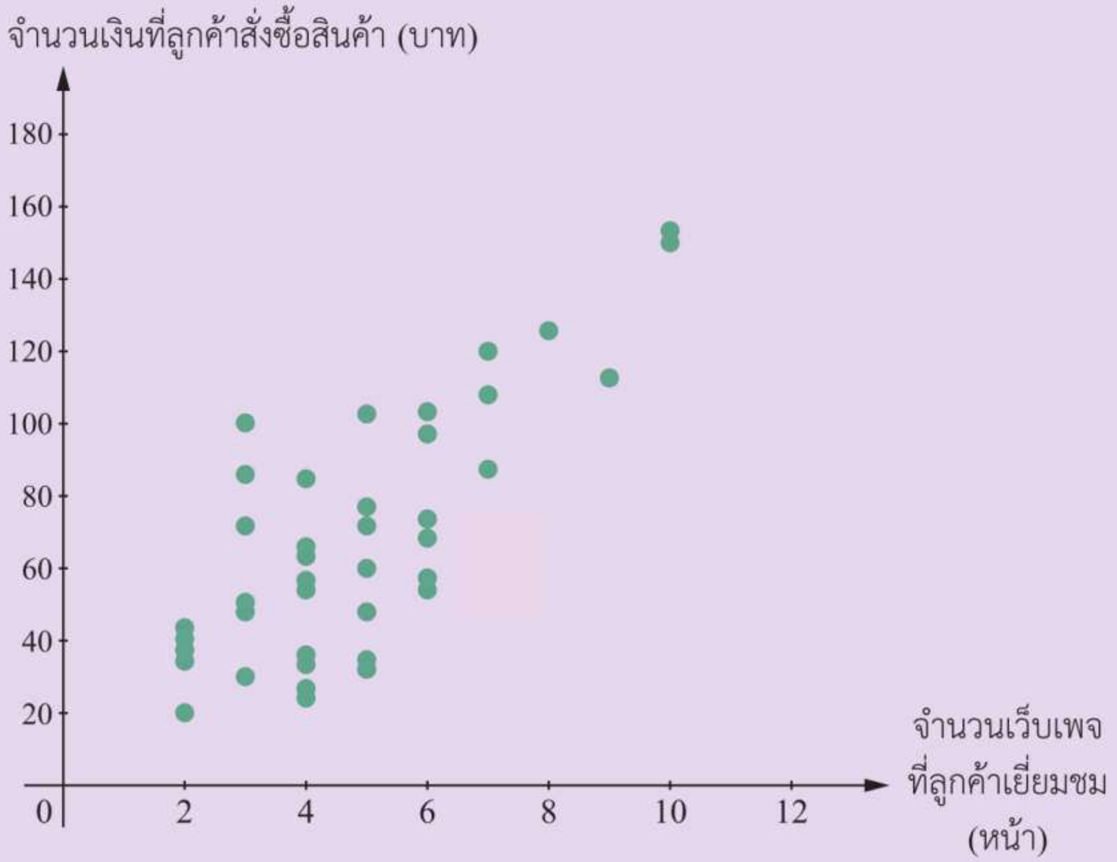
ค.


20,000
15,000
10,000
5,000
0

ง.


18,000
14,000
10,000
6,000

11 ร้านค้าออนไลน์แห่งหนึ่งต้องการศึกษาความสัมพันธ์ระหว่างจำนวนเงินที่ลูกค้าสั่งซื้อสินค้ากับจำนวนเว็บเพจที่ลูกค้าเยี่ยมชม จึงสุ่มข้อมูลการสั่งซื้อสินค้ามา 40 รายการ แล้วนำข้อมูลดังกล่าวมาเขียนแผนภาพการกระจายได้ดังนี้



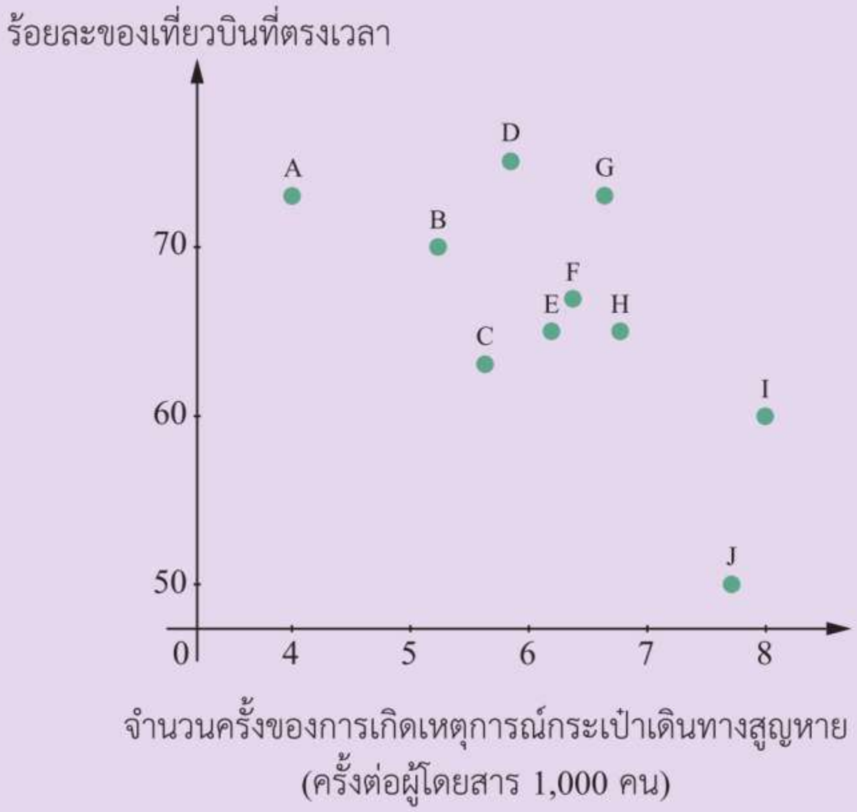
จงพิจารณาว่าจำนวนเงินที่ลูกค้าสั่งซื้อสินค้าและจำนวนเว็บเพจที่ลูกค้าเยี่ยมชมมีความสัมพันธ์กันหรือไม่ อย่างไร

12 นักวิจัยได้เก็บข้อมูลจำนวนสายพันธุ์พืชต่อพื้นที่ 0.04 ตารางเมตร ณ พื้นที่ที่มีความสูงจากระดับน้ำทะเลปานกลางต่าง ๆ กัน ได้ผลสำรวจดังนี้

ตัวอย่างที่	ความสูงจากระดับน้ำทะเลปานกลาง (กิโลเมตร)	จำนวนสายพันธุ์พืชต่อพื้นที่ 0.04 ตารางเมตร
1	0.64	15.0
2	0.64	20.0
3	0.86	16.0
4	0.86	18.5
5	0.89	13.5
6	0.89	16.0
7	1.22	11.0
8	1.22	19.5
9	1.45	11.5
10	1.45	12.0
11	1.72	8.0
12	1.72	8.5

จงเขียนแผนภาพการกระจายของข้อมูลชุดนี้ พร้อมทั้งพิจารณาว่าจำนวนสายพันธุ์พืชต่อพื้นที่ 0.04 ตารางเมตร และความสูงจากระดับน้ำทะเลปานกลางมีความสัมพันธ์กันหรือไม่ อย่างไร

13 ผลการสำรวจสายการบิน 10 แห่ง ได้แก่ สายการบิน A, B, C, D, E, F, G, H, I และ J เกี่ยวกับร้อยละของเที่ยวบินที่ตรงเวลาและจำนวนครั้งของการเกิดเหตุการณ์กระเป๋าเดินทางสูญหาย (ครั้งต่อผู้โดยสาร 1,000 คน) แสดงด้วยแผนภาพการกระจายได้ดังนี้



1) สายการบินใดตรงเวลาที่สุด และสายการบินใดไม่ตรงเวลาที่สุด
2) สายการบินใดที่เกิดเหตุการณ์กระเป๋าเดินทางสูญหายมากที่สุด และสายการบินใดที่เกิดเหตุการณ์กระเป๋าเดินทางสูญหายน้อยที่สุด
3) ข้อสรุปที่ว่า “สายการบิน J เกิดเหตุการณ์กระเป๋าเดินทางสูญหายบ่อยกว่าสายการบิน B ประมาณ 2 เท่า” เป็นจริงหรือไม่ เพราะเหตุใด
4) ข้อสรุปที่ว่า “สายการบินที่เกิดเหตุการณ์กระเป๋าเดินทางสูญหายบ่อย มีแนวโน้มที่จะตรงเวลา” เป็นจริงหรือไม่ เพราะเหตุใด

14. กำหนดข้อมูลของประชากรในแต่ละชุดดังต่อไปนี้

ชุด ก: 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 8, 9, 9, 10, 10

ชุด ข: 1, 1, 1, 1, 2, 2, 2, 3, 3, 4, 30

จงหาค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลแต่ละชุด

15. กำหนดข้อมูลของตัวอย่างในแต่ละชุดดังต่อไปนี้

ชุด ก: 2, 3, 3, 5, 5, 5, 7, 7, 8, 10, 14, 14, 19

ชุด ข: 0.9, 1.2, 1.7, 2.1, 2.5, 2.8, 3.2, 3.3, 3.7, 4.8, 5.7

ชุด ค: 59, 73, 82, 87, 87, 90

จงหา

1) ค่าเฉลี่ยเลขคณิต มัธยฐาน และฐานนิยมของข้อมูลแต่ละชุด
2) พิสัย พิสัยระหว่างควอร์ไทล์ ส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวนของข้อมูลแต่ละชุด
3) ข้อมูลชุดที่มีการกระจายมากที่สุด

16. ศิริวิทย์ได้รับผลการสอบวิชาภาษาไทย คณิตศาสตร์ ภาษาอังกฤษ และคอมพิวเตอร์ พบว่าค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาไทยเท่ากับ 21.5 คะแนน และค่าเฉลี่ยเลขคณิตของคะแนนสอบวิชาภาษาอังกฤษและวิชาคอมพิวเตอร์เท่ากับ 28.5 คะแนน จงหาค่าเฉลี่ยเลขคณิตของคะแนนสอบทั้งสี่วิชานี้

17. ถ้าค่าเฉลี่ยเลขคณิตของอายุนักเรียนชั้นมัธยมศึกษาปีที่ 3, 4 และ 5 ของโรงเรียนแห่งหนึ่งเท่ากับ 15, 17 และ 18 ปี ตามลำดับ และจำนวนนักเรียนชั้นมัธยมศึกษาปีที่ 3, 4 และ 5 ของโรงเรียนแห่งนี้เท่ากับ 60, 50 และ 40 คน ตามลำดับ จงหาค่าเฉลี่ยเลขคณิตของอายุนักเรียนชั้นมัธยมศึกษาปีที่ 3 – 5 ของโรงเรียนแห่งนี้

18 นักเรียนชั้นมัธยมศึกษาปีที่ 6 คนหนึ่งได้รับใบรายงานผลการเรียนภาคเรียนสุดท้ายดังนี้

รหัสวิชา	ชื่อวิชา	หน่วยกิต	คะแนน	เกรด
	สาระการเรียนรู้พื้นฐาน			
ท33102	ภาษาไทย 6	1.0	78	3.5
ค33102	คณิตศาสตร์ 6	1.0	70	3
ส33103	สังคมศึกษา 6	0.5	82	4
ส33104	พระพุทธศาสนา 6	0.5	85	4
ศ33101	ศิลปะ 5	0.5	74	3
ง33102	เครือข่ายและโครงงานคอมพิวเตอร์	0.5	68	2.5
อ33102	ภาษาอังกฤษ 6	1.0	87	4
	สาระการเรียนรู้เพิ่มเติม			
ท30206	เสริมทักษะภาษาไทย 2	1.0	76	3.5
ว30282	วิทยาศาสตร์เพิ่มเติม 2	0.5	63	2
พ33201	ชีวิตกับสุขภาพ 3	0.5	88	4
พ33202	กีฬากับสุขภาพ 3	0.5	90	4
ญ33202	ภาษาญี่ปุ่น 6	3.0	87	4
อ30206	ภาษาอังกฤษเพื่อการสื่อสาร 6	0.5	79	3.5
อ33205	ภาษาอังกฤษเพื่อการอ่าน – เขียน 6	1.0	84	4
อ33206	ภาษาอังกฤษรอบรู้ 6	1.0	82	4

1) จงหาเกรดเฉลี่ยของภาคเรียนสุดท้ายของนักเรียนคนนี้
2) ถ้านักเรียนคนนี้มีหน่วยกิตสะสมและเกรดเฉลี่ยของ 5 ภาคเรียนที่ผ่านมาเป็น 75.0 หน่วยกิต และ 3.75 ตามลำดับ จงหาเกรดเฉลี่ยของทั้ง 6 ภาคเรียนของนักเรียนคนนี้

19. ข้อมูลชุดหนึ่งประกอบด้วยจำนวนเต็ม 5 จำนวน โดยค่าเฉลี่ยเลขคณิตของข้อมูลชุดนี้เท่ากับ 360 เมื่อเรียงข้อมูลจากมากไปน้อย ปรากฏว่าข้อมูล 2 ตัวสุดท้าย คือ 102 และ 99 ตามลำดับ จงหาค่าที่มากที่สุดที่เป็นไปได้ของข้อมูลชุดนี้

20. ข้อมูลชุดหนึ่งประกอบด้วยจำนวนเต็ม 5 จำนวน โดยค่าเฉลี่ยเลขคณิต มัธยฐาน ฐานนิยม และพิสัยของข้อมูลชุดนี้เท่ากับ 5 จงหาชุดของข้อมูลที่เป็นไปได้ทั้งหมด

21. ร้านค้าแห่งหนึ่งต้องการวิเคราะห์ข้อมูลการขายสินค้าของร้าน จึงได้บันทึกจำนวนสินค้า (ชิ้น) ที่ขายได้ในหนึ่งวัน โดยสุ่มเก็บข้อมูลเพียง 11 วัน ในหนึ่งเดือนที่ผ่านมา ได้ข้อมูลดังนี้

85 125 30 75 80 65 90 75 78 92 67

1) จงหาควอร์ไทล์ที่ 1 ควอร์ไทล์ที่ 2 และควอร์ไทล์ที่ 3 ของข้อมูลชุดนี้
2) จงหาพิสัยและพิสัยระหว่างควอร์ไทล์ของข้อมูลชุดนี้ และพิจารณาว่าค่าวัดการกระจายสัมบูรณ์ใดเหมาะสำหรับใช้อธิบายลักษณะการกระจายของข้อมูลชุดนี้ พร้อมทั้งให้เหตุผลประกอบ
3) จงหาเปอร์เซ็นไทล์ที่ 25 เปอร์เซ็นไทล์ที่ 50 เปอร์เซ็นไทล์ที่ 75 และเปอร์เซ็นไทล์ที่ 90 ของข้อมูลชุดนี้

22. ความสูง (เซนติเมตร) ของนักกีฬาบาสเกตบอลชายและหญิงของโรงเรียนแห่งหนึ่ง ซึ่งมีจำนวนทั้งหมด 40 คน โดยเรียงข้อมูลจากน้อยไปมาก แสดงได้ดังนี้

นักกีฬาบาสเกตบอลชาย

175	176	179	180	180
180	181	183	185	188
189	190	190	195	195
195	196	198	198	206

นักกีฬาบาสเกตบอลหญิง

153	154	157	160	161
163	166	167	168	168
168	170	171	171	172
172	174	175	181	182

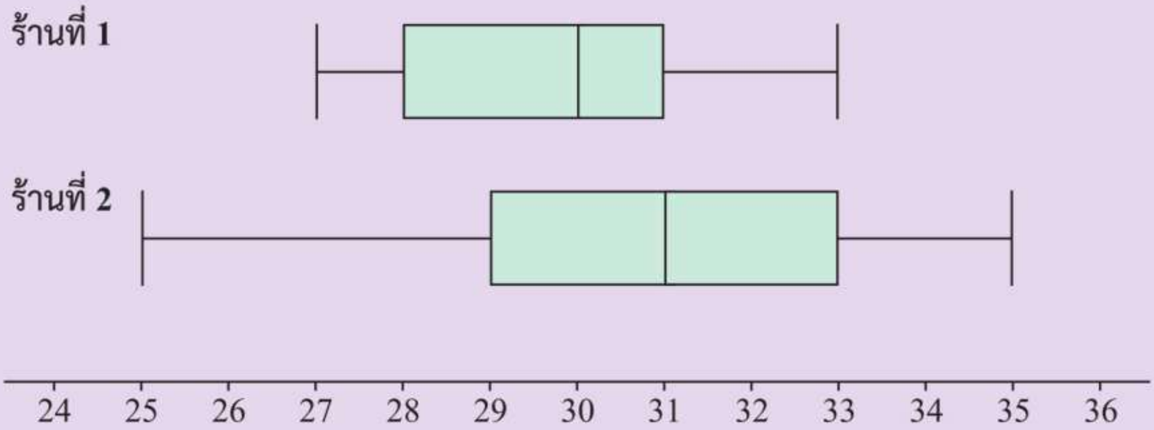
1) จงเขียนแผนภาพลำต้นและใบเพื่อนำเสนอข้อมูลสองชุดนี้ในแผนภาพเดียวกัน
2) จากแผนภาพในข้อ 1) ควรใช้ค่ากลางใดเป็นตัวแทนของข้อมูลแต่ละชุด เพราะเหตุใด
3) นักกีฬาบาสเกตบอลชายที่สูงมากกว่าค่าเฉลี่ยเลขคณิตคิดเป็นร้อยละเท่าใดของจำนวนนักกีฬาบาสเกตบอลชายทั้งหมด

23. เวลา (นาที) ที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียน 2 ห้องเรียน ที่สุ่มมาจำนวนห้องละ 10 คน แสดงได้ดังนี้

ห้อง 1	0	20	30	42	35	82	54	28	0	63
ห้อง 2	45	40	62	10	24	15	30	60	95	120

จงหาสัมประสิทธิ์การแปรผันของเวลาที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียนแต่ละห้องที่สุ่มมา พร้อมทั้งเปรียบเทียบการกระจายของเวลาที่ใช้ในการอ่านหนังสือในหนึ่งวันของนักเรียนทั้งสองห้อง

24 อรรถฤทธิ์ได้สำรวจจำนวนลูกเกดที่ใส่ในขนมปังแต่ละอันที่ซื้อจากร้านขายขนมปัง 2 แห่ง และนำเสนอด้วยแผนภาพกล่องดังนี้



1) จงหามัธยฐานของจำนวนลูกเกดที่ใส่ในขนมปังแต่ละอันของแต่ละร้าน
2) ขนมปังจากร้านที่ 1 ที่มีจำนวนลูกเกดน้อยกว่า 31 เม็ด คิดเป็นร้อยละเท่าใดของจำนวนขนมปังที่ซื้อจากร้านที่ 1 ทั้งหมด
3) จงหาควอร์ไทล์ที่ 3 ของจำนวนลูกเกดที่ใส่ในขนมปังแต่ละอันที่ซื้อจากร้านที่ 2
4) ถ้าอรรถฤทธิ์ต้องการซื้อขนมปังลูกเกดมาขาย เขาควรซื้อขนมปังลูกเกดจากร้านใดจึงจะมีโอกาสได้ลูกเกดในขนมปังแต่ละอันใกล้เคียงกัน เพราะเหตุใด
5) ถ้าอรรถฤทธิ์ชอบรับประทานลูกเกด เขาควรซื้อขนมปังจากร้านใด จึงจะมีโอกาสได้ขนมปังที่มีลูกเกดมากกว่า เพราะเหตุใด

25 มานี ชูใจ และปิติ ได้ค่าขนมจากผู้ปกครอง โดยค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของค่าขนมของทั้งสามคนเท่ากับ 50 และ 0 บาท ตามลำดับ ถ้านำค่าขนมของมานะมาคำนวณด้วยจะได้ค่าเฉลี่ยเลขคณิตใหม่เป็น 45 บาท จงหาส่วนเบี่ยงเบนมาตรฐานของค่าขนมของทั้งสี่คน

26 กำหนดให้ข้อมูลของประชากรจำนวน 2 ชุด มีลักษณะดังต่อไปนี้

ชุด ก: ประกอบด้วยจำนวนเต็มที่แตกต่างกัน 5 จำนวน โดยทั้ง 5 จำนวนนี้อยู่ระหว่าง 0 และ 10

ชุด ข: ประกอบด้วยจำนวนเต็มที่แตกต่างกัน 5 จำนวน โดยทั้ง 5 จำนวนนี้อยู่ระหว่าง 10 และ 20

1) จงยกตัวอย่างข้อมูลชุด ก และ ข ที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ก มากกว่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ข
2) มีโอกาสที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลแต่ละชุดจะเป็น 0 หรือไม่ เพราะเหตุใด
3) จงยกตัวอย่างข้อมูลชุด ก และ ข ที่ส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ก เท่ากับส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุด ข

27 กำหนดข้อมูลของประชากรชุดหนึ่งเรียงลำดับจากน้อยไปมากได้ดังนี้

$$1 \quad 2 \quad x \quad 3 \quad 3 \quad y \quad 6$$

ถ้าค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 3 และ $\frac{4\sqrt{7}}{7}$ ตามลำดับ จงหา x และ y

28. ข้อมูลของตัวอย่างชุดหนึ่งมี 20 ตัว โดยค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้เท่ากับ 10 และ 2 ตามลำดับ ถ้าภายหลังพบว่ามีการบันทึกข้อมูลตัวหนึ่งผิดพลาดไปจาก 12 บันทึกผิดเป็น 8 จงหาค่าเฉลี่ยเลขคณิตและความแปรปรวนที่ถูกต้องของข้อมูลชุดนี้

29 จงพิจารณาว่าข้อสรุปต่อไปนี้ถูกต้องหรือไม่ พร้อมทั้งให้เหตุผลประกอบ

1) ถ้าคะแนนสอบของสมชายตรงกับ P_{30} แล้วสมชายสอบได้ 30% ของคะแนนเต็ม
2) ถ้าคะแนนสอบของสมหญิงตรงกับ P_{40} แล้วจะมีนักเรียนสอบได้คะแนนมากกว่าสมหญิงอยู่ 60 คน

30 ข้อสอบวิชาหนึ่งมีคะแนนเต็ม 80 คะแนน โดยคะแนนของผู้เข้าสอบ 10 คน แสดงได้ดังนี้

25 62 59 38 38 41 69 72 26 44

1) ถ้าเกณฑ์ในการสอบผ่านคือต้องได้คะแนนไม่ต่ำกว่าร้อยละ 70 ของคะแนนเต็ม จงหาคะแนนต่ำสุดของผู้ที่สอบผ่าน
2) ถ้าเกณฑ์ในการสอบผ่านคือต้องได้คะแนนไม่ต่ำกว่าเปอร์เซ็นไทล์ที่ 70 จงหาคะแนนต่ำสุดของผู้ที่สอบผ่าน

31 ปริมาณน้ำฝน (มิลลิเมตร) ของจังหวัดในภาคเหนือ ในเดือนสิงหาคม พ.ศ. 2556 แสดงได้ดังนี้

จังหวัด	ปริมาณน้ำฝน (มิลลิเมตร)
กำแพงเพชร	214.2
เชียงใหม่	299.4
เชียงราย	423.2
ตาก	112.2
นครสวรรค์	84.0
น่าน	224.7
เพชรบูรณ์	164.0
แพร่	365.2
พะเยา	229.1
พิษณุโลก	197.5
แม่ฮ่องสอน	320.6
ลำปาง	220.9
ลำพูน	142.5
อุตรดิตถ์	392.4

จงหา

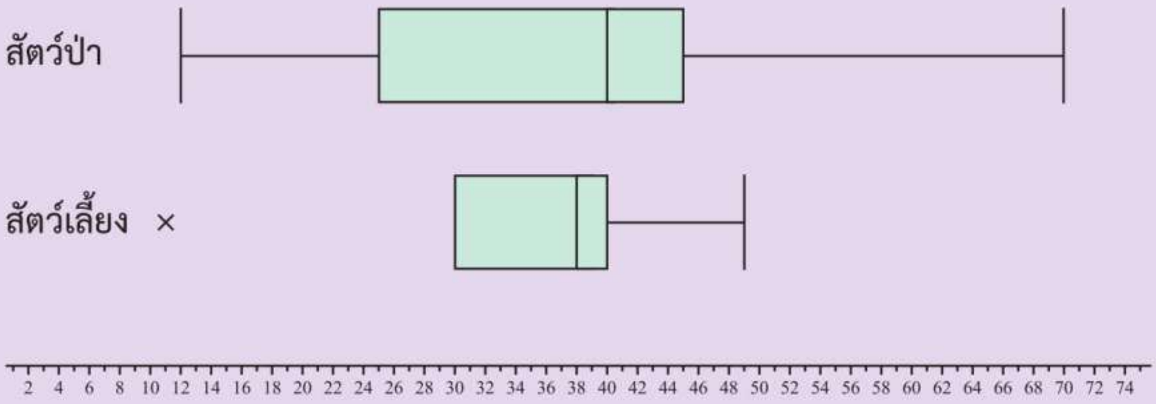
1) ปริมาณน้ำฝนที่มีจังหวัดในภาคเหนือประมาณร้อยละ 87 มีปริมาณน้ำฝนน้อยกว่า
2) ปริมาณน้ำฝนที่มีจังหวัดในภาคเหนือประมาณร้อยละ 60 มีปริมาณน้ำฝนมากกว่า

32 ผลการทดสอบความสามารถในการคิดอย่างมีวิจารณญาณของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งจำนวนทั้งหมด 40 คน ซึ่งมีคะแนนเต็ม 100 คะแนน โดยเรียงคะแนนจากน้อยไปมาก แสดงได้ดังนี้

43	45	48	49	50	51	53	54	54	54
55	56	56	58	60	60	62	63	65	65
65	66	67	69	74	75	76	76	77	78
78	80	82	84	85	92	94	96	97	98

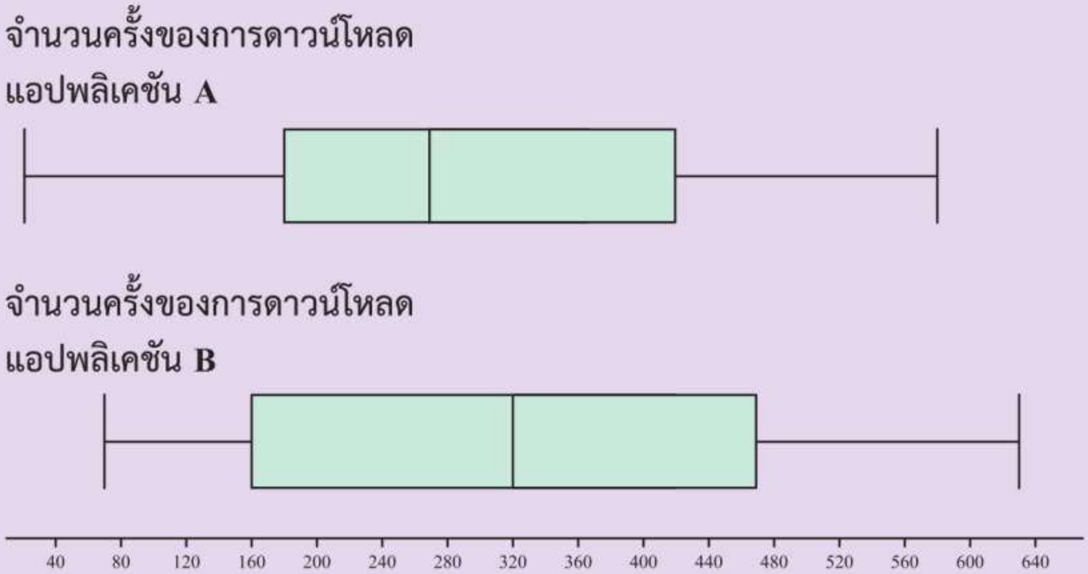
1) จงหาคะแนนที่มีนักเรียนประมาณครึ่งหนึ่งของชั้นได้คะแนนต่ำกว่า
2) จงหาคะแนนที่มีนักเรียนประมาณหนึ่งในสี่ของชั้นได้คะแนนสูงกว่า
3) จงหาคะแนนที่มีนักเรียนประมาณหกในสิบของชั้นได้คะแนนต่ำกว่า
4) ถ้าโรงเรียนแห่งนี้จัดกิจกรรมพัฒนาทักษะการคิดอย่างมีวิจารณญาณให้กับนักเรียนที่ได้คะแนนน้อยกว่าเปอร์เซ็นไทล์ที่ 25 จงหาจำนวนนักเรียนที่ต้องเข้าร่วมกิจกรรมนี้

33 อัตราเร็วสูงสุดในการเคลื่อนที่ (ไมล์ต่อชั่วโมง) ของสัตว์ป่าและสัตว์เลี้ยง แสดงด้วยแผนภาพกล่องได้ดังนี้



1) จงหามัธยฐานและค่าสูงสุดของอัตราเร็วสูงสุดในการเคลื่อนที่ของสัตว์ป่าและสัตว์เลี้ยง
2) จงหาพิสัยระหว่างควอร์ไทล์ของอัตราเร็วสูงสุดในการเคลื่อนที่ของสัตว์ป่าและสัตว์เลี้ยง
3) สามารถสรุปได้หรือไม่ว่าสัตว์ป่าส่วนใหญ่มีอัตราเร็วสูงสุดในการเคลื่อนที่น้อยกว่า 49 ไมล์ต่อชั่วโมง เพราะเหตุใด

34 จำนวนครั้งของการดาวน์โหลดแอปพลิเคชัน A และ B ตั้งแต่เดือนมกราคมถึงธันวาคม พ.ศ. 2561 แสดงด้วยแผนภาพกล่องได้ดังนี้

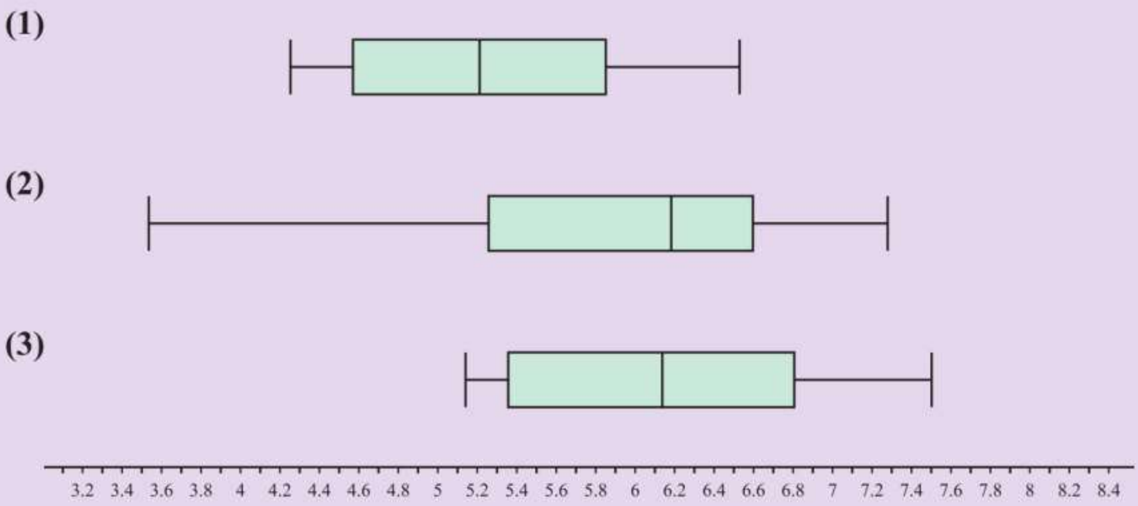


1) จงเปรียบเทียบค่าต่ำสุด มัธยฐาน และค่าสูงสุดของจำนวนครั้งของการดาวน์โหลดแอปพลิเคชันทั้งสอง
2) จงเปรียบเทียบพิสัยระหว่างควอร์ไทล์ของจำนวนครั้งของการดาวน์โหลดแอปพลิเคชันทั้งสอง

35 สหประชาชาติ (United Nations: UN) เป็นองค์การระหว่างประเทศซึ่งมีจุดมุ่งหมายในการรักษาสันติภาพและความมั่นคงระหว่างประเทศ โดยทุกปีตั้งแต่ พ.ศ. 2555 สหประชาชาติจะรายงานดัชนีความสุขของแต่ละประเทศในโลก ซึ่งมีคะแนนตั้งแต่ 0 ถึง 10 ถ้าประเทศใดมีคะแนนมากก็หมายความว่าคนในประเทศนั้นมีความสุขมาก จากรายงานใน พ.ศ. 2560 พบว่า ดัชนีความสุขของประเทศที่สุ่มมาจาก 3 ทวีป ทวีปละ 12 ประเทศ แสดงได้ดังตารางต่อไปนี้

เอเชีย		อเมริกา		ยุโรป	
เกาหลีใต้	5.838	สหรัฐอเมริกา	6.993	เบลเยียม	6.891
ไทย	6.424	แคนาดา	7.316	โครเอเชีย	5.293
จีน	5.273	เปรู	5.715	รัสเซีย	5.963
ญี่ปุ่น	5.920	ชิลี	6.652	สหราชอาณาจักร	6.714
อินเดีย	4.315	เม็กซิโก	6.578	สโลวีเนีย	5.758
เมียนมา	4.545	เวเนซุเอลา	5.250	กรีซ	5.227
อินโดนีเซีย	5.262	ปารากวัย	5.493	นอร์เวย์	7.537
ปากีสถาน	5.269	กัวเตมาลา	6.454	ฝรั่งเศส	6.442
ภูฏาน	5.011	อาร์เจนตินา	6.599	สเปน	6.403
สิงคโปร์	6.572	เฮติ	3.603	เยอรมนี	6.951
อิหร่าน	4.692	เอลซัลวาดอร์	6.003	อิตาลี	5.964
บังกลาเทศ	4.608	ฮอนดูรัส	5.181	โปรตุเกส	5.195

จากข้อมูลข้างต้น สามารถเขียนแผนภาพกล่องแสดงดัชนีความสุขของประเทศที่สุ่มมาจากแต่ละทวีปได้ดังนี้



1) จงพิจารณาว่าแผนภาพกล่อง (1), (2) และ (3) แสดงดัชนีความสุขของประเทศที่สุ่มมาจากทวีปใด
2) จงเปรียบเทียบค่าต่ำสุด มัธยฐาน และค่าสูงสุดของดัชนีความสุขของประเทศที่สุ่มมาจากแต่ละทวีป
3) เมื่อพิจารณาเฉพาะประเทศที่สุ่มมา ทวีปใดมีการกระจายของดัชนีความสุขมากที่สุด เพราะเหตุใด
4) ⍟ โดยภาพรวม เมื่อพิจารณาเฉพาะประเทศที่สุ่มมา ทวีปใดมีดัชนีความสุขมากที่สุด พร้อมทั้งให้เหตุผลประกอบ

36 เพลงรพิณได้เข้าร่วมการทดสอบวัดผลวิชาคณิตศาสตร์ระดับประเทศที่จัดโดยสถาบันแห่งหนึ่ง และได้รับใบรายงานผลสอบดังนี้

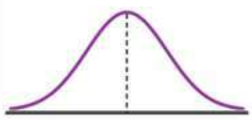
ใบรายงานผลสอบ

ชื่อ เพลงรพิณ **สกุล** อินทรเศวต **เลขประจำตัวสอบ** 001123
จำนวนผู้เข้าสอบทั้งหมด 621,519 **คะแนนที่ได้** 28 **คะแนนเต็ม** 100
คิดเป็นเปอร์เซ็นไทล์ที่ 90

ถ้าเพลงรพิณเป็นผู้เข้าสอบคนเดียวที่ได้ 28 คะแนน

1) สามารถสรุปได้หรือไม่ว่าเพลงรพิณได้คะแนนมากกว่าผู้เข้าสอบส่วนใหญ่
2) ข้อสอบที่ใช้ในการวัดผลครั้งนี้น่าจะยากหรือง่ายเกินไปสำหรับผู้เข้าสอบส่วนใหญ่ เพราะเหตุใด
3) มีผู้เข้าสอบประมาณกี่คนที่ได้คะแนนมากกว่าเพลงรพิณ
4) เป็นไปได้หรือไม่ที่คะแนนเฉลี่ยของผู้เข้าสอบทั้ง 621,519 คนนี้ มากกว่า 38 คะแนน เพราะเหตุใด

บทที่ 4 | ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น



จุดมุ่งหมาย

1. จำแนกได้ว่าตัวแปรสุ่มที่กำหนดให้เป็นตัวแปรสุ่มไม่ต่อเนื่องหรือตัวแปรสุ่มต่อเนื่อง
2. เขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง
3. หาค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มไม่ต่อเนื่อง พร้อมทั้งใช้ในการแก้ปัญหา
4. ตรวจสอบได้ว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องเป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่
5. ใช้ความรู้เกี่ยวกับการแจกแจงทวินามในการแก้ปัญหา
6. หาความน่าจะเป็นที่ตัวแปรสุ่มปกติจะมีค่าอยู่ในช่วงที่กำหนด
7. ใช้ความรู้เกี่ยวกับการแจกแจงปกติและการแจกแจงปกติมาตรฐานในการแก้ปัญหา

บทที่ 4

ตัวแปรสุ่มและการแจกแจงความน่าจะเป็น

“

ในการวางแผนและตัดสินใจทางธุรกิจ ผู้ประกอบการอาจต้องใช้ข้อมูลที่เก็บรวบรวมมาในการสร้างแบบจำลองเพื่อคาดการณ์แนวโน้มหรือความน่าจะเป็นในสถานการณ์ต่าง ๆ ตัวอย่างเช่น ผู้ประกอบการร้านขายเครื่องดื่มและของหวานอาจใช้ข้อมูลเกี่ยวกับระยะเวลาที่ลูกค้าใช้บริการที่ร้านขายเครื่องดื่มและของหวานแต่ละร้านที่สุ่มมาในละแวกนั้น ในการพิจารณาแนวโน้มหรือความน่าจะเป็นของระยะเวลาที่ลูกค้าจะเข้ามาใช้บริการที่ร้านของตน เพื่อใช้ในการวางแผนจัดเตรียมสถานที่ รวมทั้งจัดเตรียมวัตถุดิบในแต่ละช่วงเวลา เพื่อลดรายจ่ายที่ไม่จำเป็นให้ได้มากที่สุด



”

ความรู้ก่อนหน้า

- หลักการนับเบื้องต้น
- ความน่าจะเป็น
- การวิเคราะห์และนำเสนอข้อมูลเชิงปริมาณ

ipst.me/10679

สำหรับการศึกษาในบทที่ผ่านมา เป็นการวิเคราะห์ข้อมูลโดยใช้สถิติศาสตร์เชิงพรรณนา กล่าวคือ เป็นการหาข้อสรุปของข้อมูลที่เก็บมาเพื่อให้เข้าใจข้อมูลนั้นได้ง่ายขึ้น ส่วนเนื้อหาในบทนี้จะเป็นความรู้พื้นฐานสำหรับการวิเคราะห์ข้อมูลโดยใช้สถิติศาสตร์เชิงอนุมาน

เนื่องจากการวิเคราะห์ข้อมูลโดยใช้สถิติศาสตร์เชิงอนุมานเป็นการหาข้อสรุปเกี่ยวกับลักษณะของประชากรโดยใช้เพียงข้อมูลของตัวอย่างที่เก็บมา จึงทำให้สถิติศาสตร์เชิงอนุมานมีเรื่องของความไม่แน่นอนเข้ามาเกี่ยวข้อง เพราะเป็นไปได้ที่ข้อมูลของตัวอย่างจะให้ข้อสรุปทางสถิติศาสตร์ที่แตกต่างจากข้อมูลของประชากร ด้วยเหตุนี้ ความรู้เรื่องความน่าจะเป็นจึงมีบทบาทสำคัญในการศึกษาสถิติศาสตร์เชิงอนุมาน อย่างไรก็ตาม ถึงแม้จะมีความไม่แน่นอนเกิดขึ้นจากการวิเคราะห์ข้อมูลโดยใช้สถิติศาสตร์เชิงอนุมาน แต่สถิติศาสตร์เชิงอนุมานก็เป็นเรื่องสำคัญและมีประโยชน์มาก เนื่องจากโดยทั่วไปแล้ว การหาข้อมูลของประชากรเป็นไปได้ยาก เพราะไม่มีทรัพยากร งบประมาณ หรือเวลาเพียงพอในการเก็บรวบรวมข้อมูลให้ได้ครบสมบูรณ์

นอกจากนี้ การนำความน่าจะเป็นมาใช้ในการศึกษาสถิติศาสตร์ยังช่วยให้สามารถศึกษาข้อมูลที่มีจำนวนไม่แน่นอนได้ เช่น หากต้องการหาค่าเฉลี่ยของระยะเวลาที่นักท่องเที่ยวชาวต่างประเทศใช้ที่สนามบินสุวรรณภูมิ เนื่องจากนักท่องเที่ยวที่เดินทางเข้ามาที่สนามบินสุวรรณภูมิมีจำนวนไม่แน่นอน จึงไม่สามารถใช้สูตรการหาค่าเฉลี่ยเลขคณิตที่ศึกษาในบทที่ 3 ได้ แต่จะต้องนำความน่าจะเป็นมาใช้ในการศึกษาผ่านตัวแปรสุ่ม ซึ่งจะอธิบายความหมายในหัวข้อต่อไป

4.1 ความหมายและชนิดของตัวแปรสุ่ม

พิจารณาการทดลองสุ่มซึ่งได้จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง

ให้ S แทนปริภูมิตัวอย่างของการทดลองสุ่มนี้

H แทนเหรียญขึ้นหัว

และ T แทนเหรียญขึ้นก้อย

จะได้ $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

และ $n(S) = 8$

ให้ E_0, E_1, E_2 และ E_3 แทนเหตุการณ์ที่เหรียญขึ้นหัว 0, 1, 2 และ 3 ครั้ง ตามลำดับ โดยใช้ความรู้เรื่องความน่าจะเป็น จะได้ว่า

$$P(E_0) = \frac{1}{8} \qquad P(E_1) = \frac{3}{8}$$

$$P(E_2) = \frac{3}{8} \qquad P(E_3) = \frac{1}{8}$$

จากสถานการณ์ข้างต้น จะเห็นว่าสิ่งที่สนใจไม่ใช่หน้าของเหรียญที่ปรากฏในการโยนเหรียญแต่ละครั้ง แต่สนใจจำนวนครั้งที่เหรียญขึ้นหัวในการทดลองสุ่มนี้ ซึ่งมีค่าที่เป็นไปได้ 4 ค่า คือ 0, 1, 2 และ 3 โดยจะยังไม่ทราบว่าค่าที่ได้จริง ๆ คือค่าใดจนกว่าจะเสร็จสิ้นการทดลองสุ่ม จึงอาจกำหนดฟังก์ชัน X จากปริภูมิตัวอย่าง S ไปยัง $\{0, 1, 2, 3\}$ เพื่อแปลงผลลัพธ์ที่อาจเป็นไปได้ทั้งหมดของการทดลองสุ่มให้อยู่ในรูปตัวเลข โดยกำหนดให้

$$X(HHH) = 3 \qquad X(HHT) = 2$$

$$X(HTH) = 2 \qquad X(HTT) = 1$$

$$X(THH) = 2 \qquad X(THT) = 1$$

$$X(TTH) = 1 \qquad X(TTT) = 0$$

จะเรียก X ว่า **ตัวแปรสุ่ม** ซึ่งมีบทนิยามดังต่อไปนี้

ตัวแปรสุ่ม (random variable) คือฟังก์ชันจากปริภูมิตัวอย่างของการทดลองสุ่มไปยังเซตของจำนวนจริง

จะเรียกสมาชิกของเรนจ์ของตัวแปรสุ่มว่า **ค่าของตัวแปรสุ่ม** ซึ่งแต่ละค่าจะเกิดได้ด้วยความน่าจะเป็นค่าหนึ่ง

โดยทั่วไปนิยมใช้ตัวอักษรภาษาอังกฤษตัวพิมพ์ใหญ่แทนตัวแปรสุ่ม เช่น X, Y, Z และใช้ตัวอักษรภาษาอังกฤษตัวพิมพ์เล็กแทนค่าของตัวแปรสุ่ม เช่น x, y, z

จากสถานการณ์โยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง ให้ x แทนค่าที่เป็นไปได้ของตัวแปรสุ่ม X จะได้ว่า $x \in \{0, 1, 2, 3\}$

จะใช้สัญลักษณ์ $X = x$ แทนเหตุการณ์ที่เหรียญขึ้นหัว x ครั้ง

ดังนั้น จะเขียน $P(X=0)$ แทน $P(E_0)$ หรือความน่าจะเป็นที่เหรียญขึ้นหัว 0 ครั้ง
$P(X=1)$ แทน $P(E_1)$ หรือความน่าจะเป็นที่เหรียญขึ้นหัว 1 ครั้ง
$P(X=2)$ แทน $P(E_2)$ หรือความน่าจะเป็นที่เหรียญขึ้นหัว 2 ครั้ง
และ $P(X=3)$ แทน $P(E_3)$ หรือความน่าจะเป็นที่เหรียญขึ้นหัว 3 ครั้ง

โดยทั่วไป ตัวแปรสุ่มแบ่งได้เป็น 2 ชนิด ตามลักษณะของค่าที่เป็นไปได้ของตัวแปรสุ่ม ดังนี้

1. **ตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable)** คือตัวแปรสุ่มที่ค่าที่เป็นไปได้ทั้งหมดอยู่ในเซตที่สามารถนับจำนวนสมาชิกได้ หรือค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มสามารถเขียนเรียงลำดับจากน้อยไปมากได้ ทั้งนี้ เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มไม่ต่อเนื่องอาจเป็นเซตจำกัดหรือเซตอนันต์ก็ได้

   ตัวอย่างเช่น

   - ในการทอดลูกเต๋า 2 ลูก พร้อมกัน 1 ครั้ง ถ้าให้ตัวแปรสุ่มคือผลบวกของแต้มบนหน้าลูกเต๋าทั้งสอง จะได้เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มคือ $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$
   - ในการโยนเหรียญ 1 เหรียญ 1 ครั้ง ถ้าให้ตัวแปรสุ่มเป็น 0 เมื่อเหรียญขึ้นหัว และ 1 เมื่อเหรียญขึ้นก้อย จะได้เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มคือ $\{0, 1\}$
   - ในการโยนเหรียญ 1 เหรียญ ไปเรื่อย ๆ จนกว่าเหรียญจะขึ้นหัวจึงจะหยุด ถ้าให้ตัวแปรสุ่มคือจำนวนครั้งที่ต้องโยนเหรียญจนกว่าเหรียญจะขึ้นหัว จะได้เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มคือ $\{1, 2, 3, \ldots\}$ หรือ $\mathbb{N}$

2. **ตัวแปรสุ่มต่อเนื่อง (continuous random variable)** คือตัวแปรสุ่มที่เซตของค่าที่เป็นไปได้ทั้งหมดเป็นช่วงที่เป็นสับเซตของ $\mathbb{R}$

ตัวอย่างเช่น

- ให้ตัวแปรสุ่มคือความสูง (เซนติเมตร) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง อาจได้ว่าเซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเป็นช่วง $[150, 190]$
- ให้ตัวแปรสุ่มคือน้ำหนัก (กิโลกรัม) ของแตงโม ที่เก็บเกี่ยวจากสวนแห่งหนึ่ง อาจได้ว่าเซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเป็นช่วง $[1, 6]$
- ให้ตัวแปรสุ่มคือระยะเวลา (ชั่วโมง) นับจากปัจจุบันจนเกิดแผ่นดินไหวครั้งต่อไปที่จังหวัดลำปาง อาจได้ว่าเซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเป็นช่วง $[0, \infty)$

เสริมสมอง : เซตอนันต์นับได้และเซตอนันต์นับไม่ได้

เซตอนันต์สามารถแบ่งได้เป็น 2 ประเภท ได้แก่ เซตอนันต์นับได้ (countably infinite set) และเซตอนันต์นับไม่ได้ (uncountably infinite set)

เซต A ใด ๆ จะเป็นเซตอนันต์นับได้ ก็ต่อเมื่อ มีฟังก์ชันหนึ่งต่อหนึ่งจาก $\mathbb{N}$ ไปทั่วถึง A หรืออาจพิจารณาได้ว่า A เป็นเซตอนันต์นับได้ ก็ต่อเมื่อ สามารถเขียนแสดง A ได้ในรูป $A = \{a_1, a_2, a_3, ..., a_n, ...\}$ นั่นคือ สามารถเขียนแจกแจงสมาชิกทุกตัวในเซต A ในลักษณะดังกล่าวได้จนครบถ้วน โดยเขียนเพียงตัวละหนึ่งครั้งเท่านั้น ตัวอย่างของเซตอนันต์นับได้ เช่น $\mathbb{N}$, $\mathbb{Z}$ และ $\mathbb{Q}$

เรียกเซตอนันต์ซึ่งไม่ใช่เซตอนันต์นับได้ว่า เซตอนันต์นับไม่ได้ เช่น $\mathbb{Q}'$, $\mathbb{R}$ และ $\mathbb{C}$ จะเห็นว่าไม่สามารถเขียน $\mathbb{Q}'$, $\mathbb{R}$ และ $\mathbb{C}$ แบบแจกแจงสมาชิกโดยเรียงลำดับจากน้อยไปมากได้

แบบฝึกหัด 4.1

จงพิจารณาว่าตัวแปรสุ่มต่อไปนี้เป็นตัวแปรสุ่มไม่ต่อเนื่องหรือตัวแปรสุ่มต่อเนื่อง

1. ตัวแปรสุ่ม X_1 คือ จำนวนข้อสอบ (ข้อ) ที่ตอบถูก จากจำนวนข้อสอบแบบปรนัยทั้งหมด 50 ข้อ ในการสอบวิชาคณิตศาสตร์ของนักเรียนคนหนึ่ง
2. ตัวแปรสุ่ม X_2 คือ จำนวนวัยรุ่น (คน) ที่ชื่นชอบการดื่มชาเขียว จากการสอบถามวัยรุ่นจำนวน 100 คน
3. ตัวแปรสุ่ม X_3 คือ อุณหภูมิร่างกาย (องศาเซลเซียส) ของผู้ป่วยโรคไข้หวัดใหญ่ในโรงพยาบาลแห่งหนึ่ง
4. ตัวแปรสุ่ม X_4 คือ น้ำหนัก (กิโลกรัม) ของทารกแรกเกิดในโรงพยาบาลแห่งหนึ่ง
5. ตัวแปรสุ่ม X_5 คือ จำนวนลูกค้า (คน) ที่มาใช้บริการที่ธนาคารแห่งหนึ่งระหว่างเวลา 09:00 – 12:00 น.
6. ตัวแปรสุ่ม X_6 คือ เวลา (นาที) ที่ใช้ในการเดินทางจากสนามหลวงถึงอนุสาวรีย์ชัยสมรภูมิโดยรถโดยสารสาธารณะ

4.2 การแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่อง

ในการหาความน่าจะเป็นของตัวแปรสุ่มที่สนใจ นอกจากจะใช้ความรู้เรื่องความน่าจะเป็นที่นักเรียนได้ศึกษามาแล้วในชั้นมัธยมศึกษาปีที่ 5 ยังสามารถหาความน่าจะเป็นของตัวแปรสุ่มโดยใช้ความถี่สัมพัทธ์จากตารางความถี่ได้อีกด้วย ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 1

จำนวนพี่น้องของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 50 คน แสดงด้วยตารางความถี่ได้ดังนี้

จำนวนพี่น้อง (คน)	ความถี่	ความถี่สัมพัทธ์
0	6	0.12
1	22	0.44
2	17	0.34
3	4	0.08
4	1	0.02

ถ้าสุ่มนักเรียน 1 คน จากห้องนี้ และให้ตัวแปรสุ่ม X คือจำนวนพี่น้องของนักเรียนที่สุ่มได้ จงหาความน่าจะเป็นที่นักเรียนที่สุ่มได้จะมีพี่น้อง x คน เมื่อ $x \in \{0, 1, 2, 3, 4\}$

วิธีทำ สำหรับ $x \in \{0, 1, 2, 3, 4\}$ จะได้ว่า $P(X = x)$ คือความน่าจะเป็นที่นักเรียนที่สุ่มได้จะมีพี่น้อง x คน

ดังนั้น $P(X=0) = \frac{6}{50} = 0.12$

$$P(X=1) = \frac{22}{50} = 0.44$$

$$P(X=2) = \frac{17}{50} = 0.34$$

$$P(X=3) = \frac{4}{50} = 0.08$$

$$P(X=4) = \frac{1}{50} = 0.02$$

■

จากตัวอย่างข้างต้น จะเห็นว่าความน่าจะเป็นของการเกิดค่าแต่ละค่าที่เป็นไปได้ของตัวแปรสุ่มคือความถี่สัมพัทธ์

โดยทั่วไป สำหรับตัวแปรสุ่ม X ใด ๆ จะได้ $0 \le P(X = x) \le 1$ และผลรวมของความน่าจะเป็นของการเกิดค่าแต่ละค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเท่ากับ 1

เมื่อนำความน่าจะเป็นของการเกิดค่าแต่ละค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มมาเขียนแสดงเพื่ออธิบายลักษณะของตัวแปรสุ่ม จะเรียกว่า **การแจกแจงความน่าจะเป็น (probability distribution)** โดยอาจเขียนแสดงในรูปตารางหรือกราฟ เช่น จากตัวอย่างข้างต้น สามารถเขียนตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ได้ดังนี้

x	0	1	2	3	4
$P(X = x)$	0.12	0.44	0.34	0.08	0.02

ตัวอย่างที่ 2

ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่เหรียญขึ้นหัว จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตารางและกราฟ

วิธีทำ ให้ S แทนปริภูมิตัวอย่างของการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง

H แทนเหรียญขึ้นหัว

และ T แทนเหรียญขึ้นก้อย

จะได้ $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2 และ 3

เหตุการณ์ที่ $X = 0$ คือ $\{TTT\}$

เหตุการณ์ที่ $X = 1$ คือ $\{HTT, THT, TTH\}$

เหตุการณ์ที่ $X = 2$ คือ $\{HHT, HTH, THH\}$

เหตุการณ์ที่ $X = 3$ คือ $\{HHH\}$

ดังนั้น $P(X=0)=\frac{1}{8}=0.125$

$P(X=1)=\frac{3}{8}=0.375$

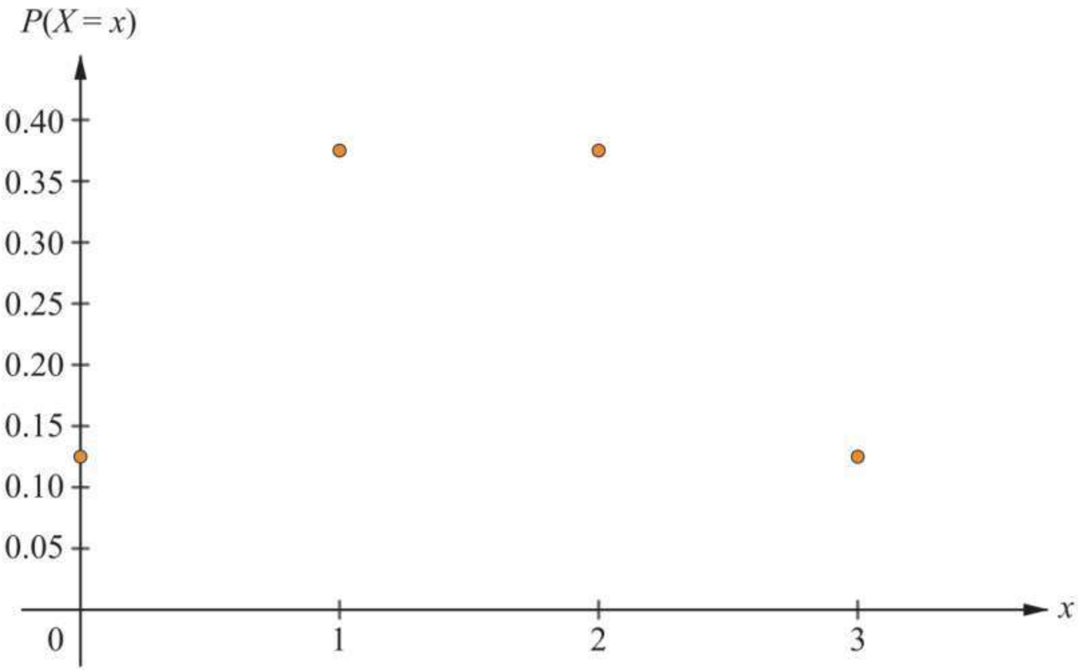
$P(X=2)=\frac{3}{8}=0.375$

$P(X=3)=\frac{1}{8}=0.125$

จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้

x	0	1	2	3
$P(X=x)$	0.125	0.375	0.375	0.125

และจะได้กราฟแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้



ตัวอย่างที่ 3

ให้ตัวแปรสุ่ม Y คือผลบวกของแต้มบนหน้าลูกเต๋า จากการทอดลูกเต๋าที่เที่ยงตรง 2 ลูก พร้อมกัน 1 ครั้ง จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y ในรูปตารางและกราฟ

วิธีทำ ให้ S แทนปริภูมิตัวอย่างของการทอดลูกเต๋าที่เที่ยงตรง 2 ลูก พร้อมกัน 1 ครั้ง

จะได้ $S = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),$

$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),$

$(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),$

$(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),$

$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),$

$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$

ค่าที่เป็นไปได้ของตัวแปรสุ่ม Y คือ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 และ 12

เหตุการณ์ที่ $Y=2$ คือ $\{(1,1)\}$

เหตุการณ์ที่ $Y=3$ คือ $\{(1,2),(2,1)\}$

เหตุการณ์ที่ $Y=4$ คือ $\{(1,3),(2,2),(3,1)\}$

เหตุการณ์ที่ $Y=5$ คือ $\{(1,4),(2,3),(3,2),(4,1)\}$

เหตุการณ์ที่ $Y=6$ คือ $\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$

เหตุการณ์ที่ $Y=7$ คือ $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

เหตุการณ์ที่ $Y=8$ คือ $\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$

เหตุการณ์ที่ $Y=9$ คือ $\{(3,6),(4,5),(5,4),(6,3)\}$

เหตุการณ์ที่ $Y=10$ คือ $\{(4,6),(5,5),(6,4)\}$

เหตุการณ์ที่ $Y=11$ คือ $\{(5,6),(6,5)\}$

เหตุการณ์ที่ $Y=12$ คือ $\{(6,6)\}$

ดังนั้น $P(Y=2)=\frac{1}{36}$ $P(Y=3)=\frac{2}{36}=\frac{1}{18}$

$P(Y=4)=\frac{3}{36}=\frac{1}{12}$ $P(Y=5)=\frac{4}{36}=\frac{1}{9}$

$P(Y=6)=\frac{5}{36}$ $P(Y=7)=\frac{6}{36}=\frac{1}{6}$

$P(Y=8)=\frac{5}{36}$ $P(Y=9)=\frac{4}{36}=\frac{1}{9}$

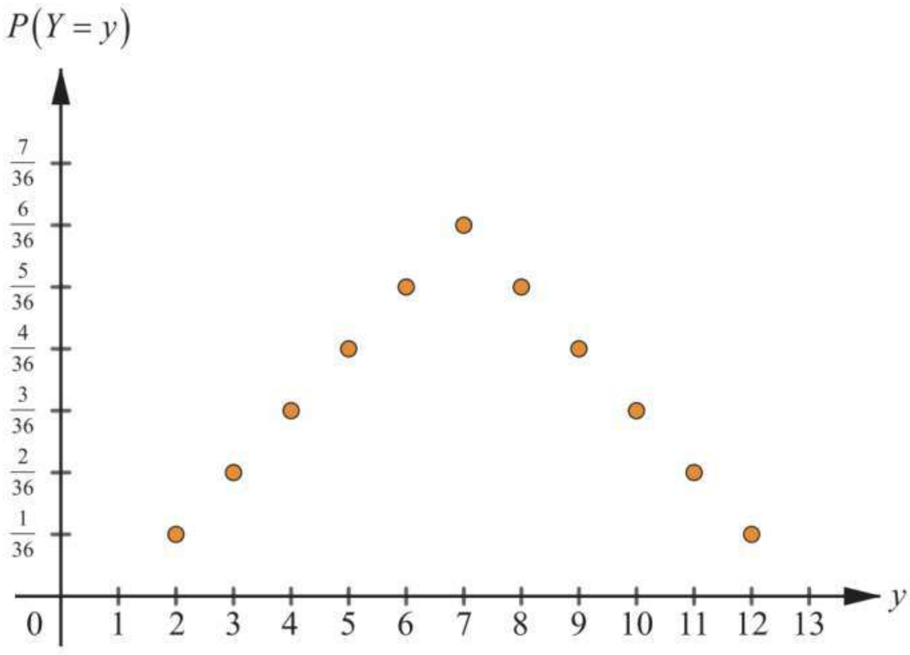
$P(Y=10)=\frac{3}{36}=\frac{1}{12}$ $P(Y=11)=\frac{2}{36}=\frac{1}{18}$

$P(Y=12)=\frac{1}{36}$

จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y ดังนี้

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

และจะได้กราฟแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y ดังนี้



แบบฝึกหัด 4.2ก

1. ข้อสอบย่อยวิชาคณิตศาสตร์ชุดหนึ่งมีทั้งหมด 10 ข้อ จำนวนข้อสอบที่ตอบถูกในการสอบย่อยวิชาคณิตศาสตร์ครั้งนี้ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่ง จำนวน 40 คน แสดงด้วยตารางความถี่ได้ดังนี้

จำนวนข้อที่ตอบถูก	0	1	2	3	4	5	6	7	8	9	10
จำนวนนักเรียน (คน)	0	1	2	5	6	3	8	7	3	3	2

ถ้าสุ่มนักเรียน 1 คน จากห้องนี้ และให้ตัวแปรสุ่ม X คือจำนวนข้อสอบที่นักเรียนคนนี้ตอบถูก จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตาราง

2. ในการตรวจสอบคุณภาพของสินค้าชนิดหนึ่งซึ่งมีทั้งหมด 80 กล่อง แต่ละกล่องมีสินค้าจำนวน 500 ชิ้น ผู้ตรวจสอบจะสุ่มสินค้าจากแต่ละกล่องมาตรวจสอบกล่องละ 100 ชิ้น โดยแสดงผลการตรวจสอบด้วยตารางความถี่ได้ดังนี้

จำนวนสินค้าที่ไม่ผ่านมาตรฐานจากสินค้า 100 ชิ้น ที่สุ่มมาตรวจสอบจากแต่ละกล่อง (ชิ้น)	0	1	2	3	4	5	6
จำนวนกล่อง	41	17	14	4	3	0	1

ถ้าสุ่มกล่องสินค้ามา 1 กล่อง จากกล่องสินค้าทั้งหมด 80 กล่อง และให้ตัวแปรสุ่ม Y คือจำนวนสินค้าที่ไม่ผ่านมาตรฐานจากสินค้า 100 ชิ้น ที่สุ่มมาตรวจสอบจากกล่องนี้ จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y ในรูปตาราง

3. ให้ตัวแปรสุ่ม Z คือผลต่างของแต้มบนหน้าลูกเต๋า จากการทอดลูกเต๋าที่เที่ยงตรง 2 ลูกพร้อมกัน 1 ครั้ง จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Z ในรูปตารางและกราฟ

ค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มไม่ต่อเนื่อง

ต่อไปจะนิยามค่าคาดหมายของตัวแปรสุ่มไม่ต่อเนื่องดังนี้

บทนิยาม 1

ค่าคาดหมาย (expected value) ของตัวแปรสุ่มไม่ต่อเนื่อง X เขียนแทนด้วย μ_X นิยามโดย

$$\mu_X = \sum_{i=1}^{n} x_i P(X = x_i)$$

เมื่อ n แทนจำนวนค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X และ $x_1, x_2, x_3, \dots, x_n$ แทนค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X

หมายเหตุ ในกรณีที่เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X เป็นเซตอนันต์ จะนิยามให้ $\mu_X = \sum_{i=1}^{\infty} x_i P(X = x_i)$ แต่ในที่นี้จะพิจารณาเฉพาะกรณีที่เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเป็นเซตจำกัด

ตัวอย่างที่ 4

จงหาค่าคาดหมายของตัวแปรสุ่ม X ที่กำหนดในตัวอย่างที่ 1

วิธีทำ จากตัวอย่างที่ 1 ตัวแปรสุ่ม X คือจำนวนพี่น้องของนักเรียนที่สุ่มได้
จะได้ว่าค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2, 3 และ 4

เนื่องจาก
$$\begin{aligned}\mu_X &= 0\cdot P(X=0)+1\cdot P(X=1)+2\cdot P(X=2)+3\cdot P(X=3) \\ &\quad +4\cdot P(X=4) \\ &= 0(0.12)+1(0.44)+2(0.34)+3(0.08)+4(0.02) \\ &= 1.44\end{aligned}$$

ดังนั้น ค่าคาดหมายของตัวแปรสุ่ม X คือ 1.44 คน ■

ในบทที่ 3 นักเรียนทราบแล้วว่าค่าเฉลี่ยเลขคณิตเป็นค่ากลางของข้อมูลที่หาได้จากการหารผลรวมของข้อมูลทั้งหมดด้วยจำนวนข้อมูลที่มี จากตัวอย่างที่ 1 จะได้ว่าค่าเฉลี่ยเลขคณิตของจำนวนพี่น้องของนักเรียน 50 คน คือ $\dfrac{0(6)+1(22)+2(17)+3(4)+4(1)}{50}=1.44$ คน

สังเกตว่า $\dfrac{0(6)+1(22)+2(17)+3(4)+4(1)}{50}=0\left(\dfrac{6}{50}\right)+1\left(\dfrac{22}{50}\right)+2\left(\dfrac{17}{50}\right)+3\left(\dfrac{4}{50}\right)+4\left(\dfrac{1}{50}\right)$

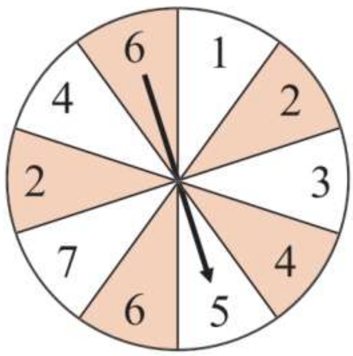
และจากตัวอย่างที่ 4 จะเห็นว่า $\mu_X=0\left(\dfrac{6}{50}\right)+1\left(\dfrac{22}{50}\right)+2\left(\dfrac{17}{50}\right)+3\left(\dfrac{4}{50}\right)+4\left(\dfrac{1}{50}\right)$

ดังนั้น ค่าเฉลี่ยเลขคณิตของจำนวนพี่น้องของนักเรียน 50 คน เท่ากับค่าคาดหมายของตัวแปรสุ่ม X เมื่อตัวแปรสุ่ม X คือจำนวนพี่น้องของนักเรียนที่สุ่มได้
จึงอาจเรียกค่าคาดหมายของตัวแปรสุ่มว่า **ค่าเฉลี่ย (mean) ของตัวแปรสุ่ม**

นอกจากจะสามารถพิจารณาค่าคาดหมายเป็นค่าเฉลี่ยของตัวแปรสุ่มแล้ว ค่าคาดหมายของตัวแปรสุ่มยังสามารถใช้ช่วยในการตัดสินใจเกี่ยวกับสถานการณ์ต่าง ๆ ได้ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 5

ในงานประจำปีของจังหวัดหนึ่งมีเกมวงล้อเสี่ยงโชค โดยมีกติกาว่า ผู้เล่นจะต้องหมุนวงล้อที่มีหมายเลข $1-7$ กำกับไว้ดังรูป ถ้าลูกศรชี้ที่ช่องที่มีหมายเลขที่กำกับเป็นจำนวนคี่ ผู้เล่นจะได้เงินรางวัล 20 บาท สมมติในการหมุนวงล้อแต่ละครั้งโอกาสที่ลูกศรจะชี้ที่ช่องใดช่องหนึ่งเท่ากัน และในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะต้องจ่ายเงินซื้อตั๋วราคา 10 บาท จงหาค่าคาดหมายของจำนวนเงินที่ผู้เล่นจะได้รับหรือเสียไป พร้อมทั้งอธิบายความหมาย



วิธีทำ ให้ตัวแปรสุ่ม X คือกำไร (ขาดทุน) ของผู้เล่นจากการหมุนวงล้อ 1 ครั้ง

เนื่องจากในการหมุนวงล้อ 1 ครั้ง มีเหตุการณ์ที่เป็นไปได้ 2 เหตุการณ์ คือ ผู้เล่นได้เงินรางวัล 20 บาท และผู้เล่นไม่ได้เงินรางวัล

แต่ในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะต้องเสียเงิน 10 บาท

จะได้ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 10 และ -10

เนื่องจากผู้เล่นจะได้รับเงินรางวัลก็ต่อเมื่อลูกศรชี้ที่ช่องที่มีหมายเลขที่กำกับเป็นจำนวนคี่ ซึ่งมี 4 ช่อง จากจำนวนช่องทั้งหมด 10 ช่อง

ดังนั้น $P(X=10)=\frac{4}{10}$ และ $P(X=-10)=\frac{6}{10}$

จะได้
$$\begin{aligned}\mu_X &= 10\left(\frac{4}{10}\right)+(-10)\left(\frac{6}{10}\right)\\ &= -2\end{aligned}$$

นั่นคือ ค่าคาดหมายของกำไร (ขาดทุน) จากการหมุนวงล้อคือ -2 บาท

ซึ่งหมายความว่า โดยเฉลี่ยแล้วในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะขาดทุนครั้งละ 2 บาท แสดงว่า ถ้าเล่นเกมวงล้อเสี่ยงโชคหลาย ๆ ครั้ง โดยเฉลี่ยแล้วผู้เล่นจะเสียเปรียบ ■

นอกจากนี้จะสามารถนิยามส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มเพื่อใช้ในการวัดการกระจายของค่าที่เป็นไปได้ของตัวแปรสุ่มว่ามีความแตกต่างจากค่าคาดหมายมากหรือน้อยเพียงใด ดังบทนิยามต่อไปนี้

บทนิยาม 2

ส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มไม่ต่อเนื่อง X เขียนแทนด้วย σ_X นิยามโดย

$$\sigma_X = \sqrt{\sum_{i=1}^{n}(x_i-\mu_X)^2 P(X=x_i)}$$

และเรียก σ_X^2 ว่า **ความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง** X

เมื่อ n แทนจำนวนค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X และ $x_1, x_2, x_3, \ldots, x_n$ แทนค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X

หมายเหตุ ในกรณีที่เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม X เป็นเซตอนันต์ จะนิยามให้ $\sigma_X = \sqrt{\sum_{i=1}^{\infty}(x_i - \mu_X)^2 P(X = x_i)}$ แต่ในที่นี้จะพิจารณาเฉพาะกรณีที่เซตของค่าที่เป็นไปได้ทั้งหมดของตัวแปรสุ่มเป็นเซตจำกัด

ตัวอย่างที่ 6

ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่เหรียญขึ้นหัว จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

วิธีทำ จากตัวอย่างที่ 2 จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้

x	0	1	2	3
$P(X = x)$	0.125	0.375	0.375	0.125

เนื่องจาก $\mu_X = 0(0.125) + 1(0.375) + 2(0.375) + 3(0.125)$

$= 1.5$

ดังนั้น ค่าคาดหมายของตัวแปรสุ่ม X คือ 1.5 ครั้ง

เนื่องจาก $\sigma_X^2 = (0-1.5)^2(0.125) + (1-1.5)^2(0.375) + (2-1.5)^2(0.375) + (3-1.5)^2(0.125)$

$= 0.75$

ดังนั้น ความแปรปรวนของตัวแปรสุ่ม X คือ 0.75 ครั้ง^2

เนื่องจาก $\sigma_X = \sqrt{0.75}$

≈ 0.87

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X มีค่าประมาณ 0.87 ครั้ง ■

ตัวอย่างที่ 7

ให้ตัวแปรสุ่ม Y คือผลบวกของแต้มบนหน้าลูกเต๋า จากการทอดลูกเต๋าที่เที่ยงตรง 2 ลูก พร้อมกัน 1 ครั้ง จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม Y

วิธีทำ จากตัวอย่างที่ 3 จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y ดังนี้

y	2	3	4	5	6	7	8	9	10	11	12
$P(Y=y)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

เนื่องจาก
$$\begin{aligned}\mu_Y &= 2\left(\frac{1}{36}\right)+3\left(\frac{1}{18}\right)+4\left(\frac{1}{12}\right)+5\left(\frac{1}{9}\right)+6\left(\frac{5}{36}\right)+7\left(\frac{1}{6}\right)+8\left(\frac{5}{36}\right)\\ &\quad +9\left(\frac{1}{9}\right)+10\left(\frac{1}{12}\right)+11\left(\frac{1}{18}\right)+12\left(\frac{1}{36}\right)\\ &= 7\end{aligned}$$

ดังนั้น ค่าคาดหมายของตัวแปรสุ่ม Y คือ 7 แต้ม

เนื่องจาก
$$\begin{aligned}\sigma_Y^2 &= (2-7)^2\left(\frac{1}{36}\right)+(3-7)^2\left(\frac{1}{18}\right)+(4-7)^2\left(\frac{1}{12}\right)+(5-7)^2\left(\frac{1}{9}\right)\\ &\quad +(6-7)^2\left(\frac{5}{36}\right)+(7-7)^2\left(\frac{1}{6}\right)+(8-7)^2\left(\frac{5}{36}\right)+(9-7)^2\left(\frac{1}{9}\right)\\ &\quad +(10-7)^2\left(\frac{1}{12}\right)+(11-7)^2\left(\frac{1}{18}\right)+(12-7)^2\left(\frac{1}{36}\right)\\ &= \frac{35}{6}\\ &\approx 5.83\end{aligned}$$

ดังนั้น ความแปรปรวนของตัวแปรสุ่ม Y มีค่าประมาณ 5.83 แต้ม^2

เนื่องจาก
$$\begin{aligned}\sigma_Y &= \sqrt{\frac{35}{6}}\\ &\approx 2.42\end{aligned}$$

ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม Y มีค่าประมาณ 2.42 แต้ม ■

แบบฝึกหัด 4.2ข

1. จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มในข้อ 1 – 3 ของแบบฝึกหัด 4.2ก

2. กล่องใบหนึ่งบรรจุเบี้ย 6 อัน โดยมีหมายเลข 3, 5, 6, 7, 8 และ 11 กำกับไว้ ถ้าสุ่มหยิบเบี้ย 2 อัน โดยหยิบเบี้ยทีละอันและไม่ใส่คืนก่อนหยิบเบี้ยอันที่สอง และให้ตัวแปรสุ่ม X คือผลบวกของหมายเลขบนเบี้ยทั้งสองอันที่สุ่มได้
   1) จงหาค่าคาดหมายของตัวแปรสุ่ม X พร้อมทั้งอธิบายความหมาย
   2) จงหาความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

3. ลูกค้ารายหนึ่งต้องการทำประกันชีวิตกับบริษัทมั่นใจประกันชีวิต โดยกำหนดทุนประกัน 2,000,000 บาท (นั่นคือ ในกรณีที่ลูกค้าเสียชีวิต บริษัทจะต้องจ่ายเงินให้ผู้รับประโยชน์ที่ลูกค้าระบุไว้เป็นจำนวนเงิน 2,000,000 บาท) และลูกค้าจะต้องจ่ายค่าเบี้ยประกันปีละ 50,000 บาท ถ้าลูกค้ารายนี้มีภาวะหยุดหายใจขณะหลับ (obstructive sleep apnea) โดยโอกาสที่เขาจะเสียชีวิตในแต่ละปีคิดเป็นร้อยละ 1 จงพิจารณาว่าถ้าบริษัทมั่นใจประกันชีวิตรับทำประกันชีวิตให้กับลูกค้ารายนี้ บริษัทจะได้กำไรหรือขาดทุนโดยเฉลี่ยปีละกี่บาท

4. จากการตรวจสอบคุณภาพของสินค้าชนิดหนึ่งจำนวน 60 ชิ้น ได้ข้อมูลจำนวนรอยตำหนิบนสินค้าแต่ละชิ้น ดังตารางต่อไปนี้

จำนวนรอยตำหนิบนสินค้า (แห่ง)	0	1	2	3
จำนวนสินค้า (ชิ้น)	47	4	6	3

ถ้าผู้ขายได้กำไรจากการขายสินค้าที่ไม่มีรอยตำหนิชิ้นละ 20 บาท แต่ถ้าสินค้ามีรอยตำหนิจะเสียค่าซ่อมแซมแห่งละ 11 บาท จงหาค่าคาดหมายของรายได้จากการขายสินค้า 1 ชิ้น

4.2.1 การแจกแจงเอกรูปไม่ต่อเนื่อง

บทนิยาม 3

ให้ X เป็นตัวแปรสุ่มไม่ต่อเนื่อง ถ้าค่าที่เป็นไปได้ทั้งหมดของ X คือ $x_1, x_2, x_3, \dots, x_n$ แล้ว การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็น **การแจกแจงเอกรูปไม่ต่อเนื่อง (discrete uniform distribution)** เมื่อ $P(X = x_i) = \frac{1}{n}$ สำหรับทุก $i \in \{1, 2, 3, \dots, n\}$

จากบทนิยาม จะเห็นว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่มจะเป็นการแจกแจงเอกรูปไม่ต่อเนื่อง เมื่อการเกิดค่าแต่ละค่าที่เป็นไปได้ของตัวแปรสุ่มมีความน่าจะเป็นเท่ากัน

ตัวอย่างที่ 8

ในการทอดลูกเต๋าที่เที่ยงตรง 1 ลูก 1 ครั้ง ให้ตัวแปรสุ่ม X คือแต้มบนหน้าลูกเต๋า จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่ พร้อมทั้งเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตารางและกราฟ

วิธีทำ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 1, 2, 3, 4, 5 และ 6

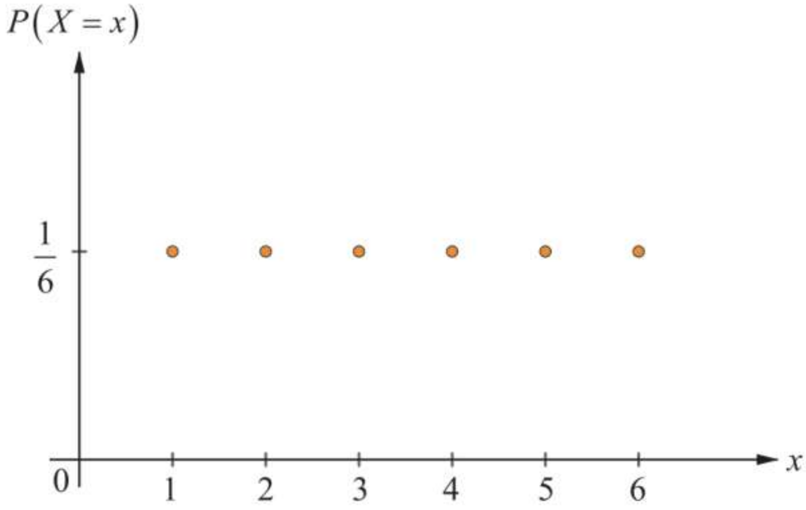
สำหรับ $x \in \{1, 2, 3, 4, 5, 6\}$ จะได้ว่า $P(X = x)$ คือความน่าจะเป็นที่ลูกเต๋าขึ้นแต้ม x

เนื่องจากลูกเต๋าเที่ยงตรง จะได้ว่า $P(X = x) = \frac{1}{6}$

ดังนั้น การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่อง จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้

x	1	2	3	4	5	6
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

และจะได้กราฟแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้



ตัวอย่างที่ 9

ลูกคิดสุ่มหยิบสลาก 1 ใบ จากกล่องที่บรรจุสลาก 4 ใบ แต่ละใบระบุจำนวนเงินรางวัลแตกต่างกันคือ 20, 50, 100 และ 500 บาท ให้ตัวแปรสุ่ม X คือจำนวนเงินรางวัลที่ลูกคิดจะได้รับ

1) จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่
2) จงหาค่าคาดหมายของตัวแปรสุ่ม X
3) ถ้าลูกคิดต้องจ่ายเงินซื้อตั๋วราคา 150 บาท เพื่อหยิบสลาก 1 ใบ จงพิจารณาว่าถ้าลูกคิดสุ่มหยิบสลากหลาย ๆ ครั้ง โดยเฉลี่ยแล้วลูกคิดได้เปรียบหรือเสียเปรียบ

วิธีทำ 1) เนื่องจากตัวแปรสุ่ม X คือจำนวนเงินที่ลูกคิดจะได้รับ

จะได้ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 20, 50, 100 และ 500

สำหรับ $x \in \{20, 50, 100, 500\}$ จะได้ว่า $P(X = x)$ คือความน่าจะเป็นที่ลูกคิดจะได้เงินรางวัล x บาท นั่นคือ $P(X = x) = \frac{1}{4}$

ดังนั้น การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่อง

2) เนื่องจาก $\mu_X = 20\left(\frac{1}{4}\right)+50\left(\frac{1}{4}\right)+100\left(\frac{1}{4}\right)+500\left(\frac{1}{4}\right)$

$= 167.50$

ดังนั้น ค่าคาดหมายของตัวแปรสุ่ม X คือ 167.50 บาท

3) ในแต่ละครั้งที่ลูกคิดซื้อตั๋วราคา 150 บาท เพื่อสุ่มหยิบสลาก มีค่าคาดหมายที่ลูกคิดจะได้เงินรางวัล 167.50 บาท ซึ่งได้เปรียบอยู่ $167.50 - 150 = 17.50$ บาท แสดงว่า ถ้าลูกคิดสุ่มหยิบสลากหลาย ๆ ครั้ง โดยเฉลี่ยแล้วลูกคิดจะได้เปรียบ

แบบฝึกหัด 4.2.1

1. จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อไปนี้เป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่ พร้อมทั้งให้เหตุผลประกอบ
   1) ตัวแปรสุ่ม X_1 คือจำนวนครั้งที่เหรียญขึ้นหัว จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 1 ครั้ง
   2) ตัวแปรสุ่ม X_2 คือจำนวนครั้งที่เหรียญขึ้นก้อย จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 10 ครั้ง
   3) ตัวแปรสุ่ม X_3 คือผลรวมของเงินรางวัลที่ได้รับ จากการสุ่มหยิบสลาก 2 ใบพร้อมกัน จากกล่องที่บรรจุสลาก 4 ใบ โดยแต่ละใบระบุจำนวนเงินรางวัลแตกต่างกันคือ 10, 30, 60 และ 80 บาท
   4) ตัวแปรสุ่ม X_4 คือจำนวนสินค้าที่ไม่ผ่านมาตรฐาน เมื่อสุ่มกล่องสินค้ามา 1 กล่อง จากกล่องสินค้าทั้งหมด 80 กล่อง โดยข้อมูลจำนวนสินค้าที่ไม่ผ่านมาตรฐานในแต่ละกล่องแสดงด้วยตารางความถี่ได้ดังนี้

จำนวนสินค้าที่ไม่ผ่านมาตรฐานในแต่ละกล่อง (ชิ้น)	0	1	2	3	4
จำนวนกล่อง	16	16	16	16	16

2. ถ้าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่อง และค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 5, 6, 7, 8, 9 และ 10 จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

3. ครูประจำชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่งได้นำของขวัญมาร่วมจับสลากในวันขึ้นปีใหม่ โดยมูลค่าของของขวัญที่น้อยที่สุดคือ 100 บาท และเมื่อนำมูลค่าของของขวัญทั้งหมดมาเรียงลำดับจากน้อยไปมาก พบว่า มูลค่าของของขวัญที่อยู่ติดกันจะต่างกัน 100 บาท ถ้าความน่าจะเป็นที่นักเรียนจะได้ของขวัญแต่ละชิ้นมีค่าเท่ากันคือ 0.125 จงหาว่าของขวัญทั้งหมดที่ครูนำมาจับสลากมีมูลค่าเท่าใด

4. ในงานประจำปีของโรงเรียนแห่งหนึ่งมีเกมลูกเต๋าเสี่ยงโชค โดยมีกติกาว่า ผู้เล่นจะต้องทอดลูกเต๋าที่เที่ยงตรง 1 ลูก ถ้าลูกเต๋าขึ้นแต้ม 1, 2, 3, 4, 5 หรือ 6 ผู้เล่นจะได้เงินรางวัล 30, 18, 12, 10, 5 หรือ a บาท ตามลำดับ โดยที่ $0 \le a \le 5$ ให้ตัวแปรสุ่ม Z คือเงินรางวัลที่ผู้เล่นจะได้รับจากการเล่นเกมลูกเต๋าเสี่ยงโชค
   1) ถ้า $a = 5$ จงพิจารณาว่าตัวแปรสุ่ม Z เป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่ พร้อมทั้งให้เหตุผลประกอบ
   2) จงยกตัวอย่างค่าของ a ที่ทำให้ตัวแปรสุ่ม Z เป็นการแจกแจงเอกรูปไม่ต่อเนื่อง
   3) ถ้า $a = 1$ ผู้จัดเกมนี้ควรตั้งราคาตั๋วสำหรับเล่นเกมอย่างน้อยเท่าใด (ตอบเป็นจำนวนเต็ม) โดยเฉลี่ยแล้วจึงจะไม่ขาดทุน เมื่อมีการเล่นเกมลูกเต๋าเสี่ยงโชคหลาย ๆ ครั้ง

5. เกมวงล้อเสี่ยงโชคมีกติกาการเล่นคือ ผู้เล่นจะต้องหมุนวงล้อรูปวงกลมที่แบ่งเป็น 10 ช่องเท่า ๆ กัน โดยแต่ละช่องระบุจำนวนเงินรางวัลแตกต่างกันคือ 50, 100, 150, 200, … , 500 บาท ถ้าลูกศรชี้ที่ช่องใด ผู้เล่นจะได้เงินรางวัลตามที่ระบุในช่องนั้น สมมติว่าในการหมุนวงล้อแต่ละครั้งโอกาสที่ลูกศรจะชี้ที่ช่องใดช่องหนึ่งเท่ากัน และในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะต้องจ่ายเงินซื้อตั๋วราคา 300 บาท นักเรียนจะเล่นเกมนี้หรือไม่ เพราะเหตุใด

4.2.2 การแจกแจงทวินาม

ในหัวข้อ 4.2.1 นักเรียนได้ศึกษาเกี่ยวกับการแจกแจงเอกรูปไม่ต่อเนื่อง ซึ่งเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มอย่างง่าย เนื่องจากการเกิดค่าแต่ละค่าที่เป็นไปได้ของตัวแปรสุ่มมีความน่าจะเป็นเท่ากัน ซึ่งอาจพบได้ไม่มากนักในชีวิตจริง สำหรับหัวข้อนี้ จะศึกษาเกี่ยวกับการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีลักษณะเฉพาะประเภทหนึ่ง ซึ่งความน่าจะเป็นของการเกิดค่าแต่ละค่าที่เป็นไปได้ของตัวแปรสุ่มไม่จำเป็นต้องเท่ากัน ดังเช่นในตัวอย่างที่ 2 เมื่อกำหนดให้ตัวแปรสุ่ม X คือจำนวนครั้งที่เหรียญขึ้นหัว จากการโยนเหรียญที่เที่ยงตรง 1 เหรียญ 3 ครั้ง จะได้ว่า X เป็นตัวแปรสุ่มไม่ต่อเนื่อง สังเกตว่าในแต่ละครั้งที่โยนเหรียญจะมีผลลัพธ์ที่เป็นไปได้ 2 แบบ คือ เหรียญขึ้นหัวหรือก้อย แต่จากการกำหนดตัวแปรสุ่ม X จะเห็นว่าสนใจจำนวนครั้งที่เหรียญขึ้นหัว จึงอาจพิจารณาว่าในการโยนเหรียญแต่ละครั้ง ถ้าเหรียญขึ้นหัวคือ **สำเร็จ** แต่ถ้าเหรียญขึ้นก้อยคือ **ไม่สำเร็จ** ดังนั้น สามารถพิจารณาว่าตัวแปรสุ่ม X คือ จำนวนครั้งของการเกิดผลสำเร็จจากการโยนเหรียญที่เที่ยงตรง 1 เหรียญเป็นจำนวน 3 ครั้ง เช่น เหตุการณ์ที่ $X = 1$ ซึ่งคือ $\{HTT, THT, TTH\}$ หมายความว่า แต่ละสมาชิกในเหตุการณ์นี้เกิดผลสำเร็จ 1 ครั้ง (เหรียญขึ้นหัว 1 ครั้ง) และไม่เกิดผลสำเร็จ 2 ครั้ง (เหรียญขึ้นก้อย 2 ครั้ง) นอกจากนี้จะเห็นว่าผลที่ได้จากการโยนเหรียญในครั้งก่อนหน้าไม่ส่งผลต่อการโยนเหรียญครั้งต่อไปและความน่าจะเป็นที่จะเกิดผลสำเร็จในการโยนเหรียญแต่ละครั้งเท่ากัน คือ $\frac{1}{2}$

การแจกแจงความน่าจะเป็นของตัวแปรสุ่มในตัวอย่างข้างต้น เรียกว่า การแจกแจงทวินาม ซึ่งมีบทนิยามดังนี้

บทนิยาม 4

การแจกแจงทวินาม (binomial distribution) คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ซึ่งคือจำนวนครั้งของการเกิดผลสำเร็จจากการทดลองสุ่ม n ครั้ง ที่เป็นอิสระกัน โดยในแต่ละครั้งมีโอกาสเกิดผลสำเร็จด้วยความน่าจะเป็นเท่ากับ p และไม่เกิดผลสำเร็จด้วยความน่าจะเป็นเท่ากับ $1-p$

หมายเหตุ 1. เรียก n และ p ว่า **พารามิเตอร์ของการแจกแจงทวินาม** และเขียนสัญลักษณ์ $X \sim B(n, p)$ เพื่อแสดงว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินามที่มี n และ p เป็นพารามิเตอร์

2. การทดลองสุ่ม 1 ครั้ง ที่มีผลลัพธ์ที่เป็นไปได้ 2 แบบ คือ สำเร็จหรือไม่สำเร็จ เรียกว่า **การลองแบร์นูลลี (Bernoulli trial)** เช่น การโยนเหรียญ 1 เหรียญ 1 ครั้ง

จากบทนิยามข้างต้น สรุปได้ว่า การแจกแจงทวินามคือการแจกแจงความน่าจะเป็นของตัวแปรสุ่มไม่ต่อเนื่องที่มีลักษณะดังต่อไปนี้

1. เกิดจากการทดลองสุ่มจำนวน n ครั้งที่เป็นอิสระกัน กล่าวคือ ผลที่ได้จากการทดลองสุ่มในครั้งก่อนหน้าไม่ส่งผลต่อการทดลองสุ่มในครั้งต่อ ๆ ไป
2. การทดลองสุ่มแต่ละครั้งมีผลลัพธ์ที่เป็นไปได้เพียง 2 แบบ คือ สำเร็จหรือไม่สำเร็จ
3. ความน่าจะเป็นที่จะเกิดผลสำเร็จในการทดลองสุ่มแต่ละครั้งเท่ากัน ให้เป็น p เมื่อ $0 < p < 1$ และจะได้ว่าความน่าจะเป็นที่จะไม่เกิดผลสำเร็จในการทดลองสุ่มแต่ละครั้งเป็น $1 - p$

ทฤษฎีบทต่อไปนี้ใช้ในการหาความน่าจะเป็น ค่าคาดหมาย และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นเป็นการแจกแจงทวินาม โดยในที่นี้จะขอละการพิสูจน์

ทฤษฎีบท 1

ถ้าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินาม จะได้ว่า

1. $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ สำหรับทุก $x \in \{0, 1, 2, \ldots, n\}$
2. $\mu_X = np$
3. $\sigma_X = \sqrt{np(1-p)}$

เมื่อ n แทนจำนวนครั้งของการทดลองสุ่ม และ p แทนความน่าจะเป็นที่จะเกิดผลสำเร็จในการทดลองสุ่มแต่ละครั้ง

ข้อสังเกต จากทฤษฎีบท 1 ข้อ 1 และทฤษฎีบททวินาม จะได้ว่า

$$\sum_{x=0}^{n} P(X=x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p+(1-p))^n = 1$$

ตัวอย่างที่ 10

ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่ลูกเต๋าขึ้นแต้ม 5 จากการทอดลูกเต๋าที่เที่ยงตรง 1 ลูก 7 ครั้ง

1) จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินามหรือไม่
2) จงหาความน่าจะเป็นที่ลูกเต๋าขึ้นแต้ม 5 เป็นจำนวน 5 ครั้ง
3) จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตารางและกราฟ
4) จงหาค่าคาดหมายและความแปรปรวนของตัวแปรสุ่ม X

วิธีทำ 1) เนื่องจากตัวแปรสุ่ม X มีลักษณะดังต่อไปนี้

1. เกิดจากการทดลองสุ่ม (การทอดลูกเต๋าที่เที่ยงตรง 1 ลูก) จำนวน 7 ครั้ง ที่เป็นอิสระกัน
2. การทดลองสุ่มแต่ละครั้งเกิดผลลัพธ์ได้เพียง 2 แบบ คือ สำเร็จ (ลูกเต๋าขึ้นแต้ม 5) หรือไม่สำเร็จ (ลูกเต๋าไม่ขึ้นแต้ม 5)
3. ความน่าจะเป็นที่ลูกเต๋าขึ้นแต้ม 5 ในการทอดลูกเต๋าที่เที่ยงตรงแต่ละครั้งเท่ากัน โดยเท่ากับ $\frac{1}{6}$ และความน่าจะเป็นที่ลูกเต๋าไม่ขึ้นแต้ม 5 ในการทอดลูกเต๋าที่เที่ยงตรงแต่ละครั้งเป็น $1-\frac{1}{6}=\frac{5}{6}$

ดังนั้น การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินาม

2) ความน่าจะเป็นที่ลูกเต๋าขึ้นแต้ม 5 เป็นจำนวน 5 ครั้ง คือ

$$P(X=5) = \binom{7}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^2 \approx 0.001875$$

3) ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2, 3, 4, 5, 6 และ 7

โดย $P(X=0)=\binom{7}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^7$ $\quad P(X=1)=\binom{7}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^6$

$P(X=2)=\binom{7}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^5$ $\quad P(X=3)=\binom{7}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^4$

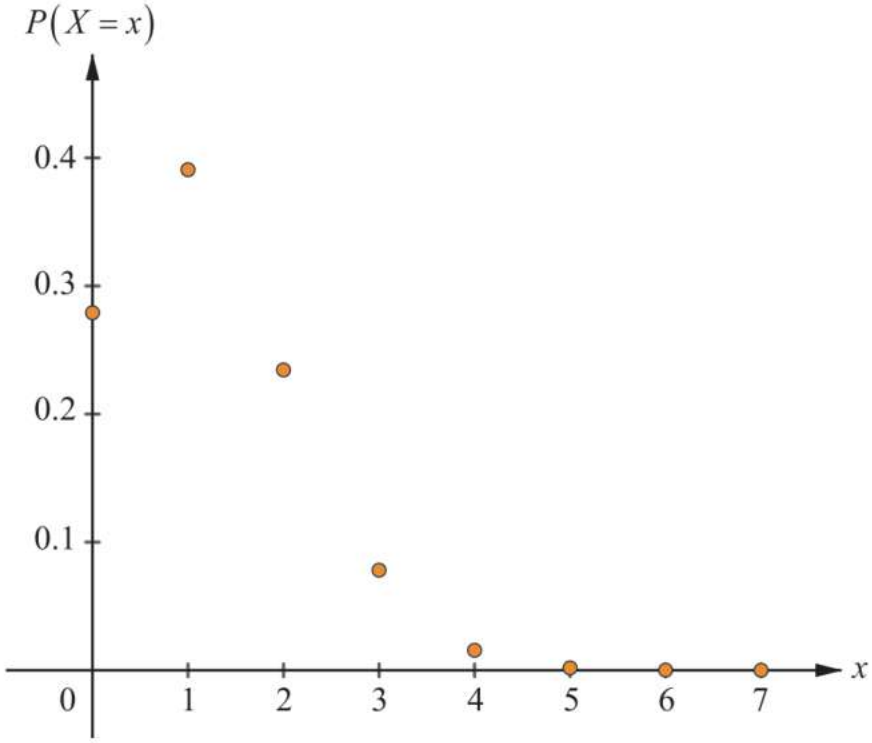
$P(X=4)=\binom{7}{4}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^3$ $\quad P(X=5)=\binom{7}{5}\left(\frac{1}{6}\right)^5\left(\frac{5}{6}\right)^2$

$P(X=6)=\binom{7}{6}\left(\frac{1}{6}\right)^6\left(\frac{5}{6}\right)^1$ $\quad P(X=7)=\binom{7}{7}\left(\frac{1}{6}\right)^7\left(\frac{5}{6}\right)^0$

จะได้ตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้

x	$P(X=x)$
0	0.279082
1	0.390714
2	0.234429
3	0.078143
4	0.015628
5	0.001875
6	0.000125
7	0.000004

และจะได้กราฟแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ดังนี้



4) เนื่องจาก $\mu_X = 7\left(\frac{1}{6}\right) \approx 1.17$

ดังนั้น ค่าคาดหมายของตัวแปรสุ่ม X มีค่าประมาณ 1.17 ครั้ง

เนื่องจาก $\sigma_X^2 = 7\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) \approx 0.97$

ดังนั้น ความแปรปรวนของตัวแปรสุ่ม X มีค่าประมาณ 0.97 ครั้ง^2

ตัวอย่างที่ 11

ในการรักษาโรคมะเร็งด้วยสมุนไพรที่คิดค้นขึ้นมาใหม่ พบว่า เมื่อผู้ป่วยรับประทานสมุนไพรชนิดนี้ต่อเนื่องกันไปตามแพทย์สั่งในช่วงระยะเวลาหนึ่ง ความน่าจะเป็นที่ผู้ป่วยแต่ละคนจะหายจากโรคมะเร็งเป็น 0.5 ถ้านักวิจัยสุ่มผู้ป่วยโรคมะเร็งที่มารับการรักษาด้วยสมุนไพรนี้จำนวน 6 คน

1) จงหาความน่าจะเป็นที่จะมีผู้ป่วยหายจากโรคมะเร็งอย่างน้อย 3 คน
2) จงหาค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของจำนวนผู้ป่วยที่หายจากโรคมะเร็ง พร้อมทั้งอธิบายความหมาย

วิธีทำ ให้ตัวแปรสุ่ม X คือจำนวนผู้ป่วยที่หายจากโรคมะเร็งจากผู้ป่วยโรคมะเร็งที่มารับการรักษาด้วยสมุนไพรนี้ที่สุ่มมาจำนวน 6 คน

จะได้ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2, 3, 4, 5 และ 6

เนื่องจากตัวแปรสุ่ม X มีลักษณะดังต่อไปนี้

1. เกิดจากการสุ่มผู้ป่วยโรคมะเร็งที่มารับการรักษาด้วยสมุนไพรนี้จำนวน 6 คนที่เป็นอิสระกัน
2. การทดลองสุ่มแต่ละครั้งเกิดผลลัพธ์ได้เพียง 2 แบบ คือ สำเร็จ (ผู้ป่วยที่สุ่มมาหายจากโรคมะเร็ง) หรือไม่สำเร็จ (ผู้ป่วยที่สุ่มมาไม่หายจากโรคมะเร็ง)
3. ความน่าจะเป็นที่ผู้ป่วยแต่ละคนจะหายจากโรคมะเร็งเท่ากัน โดยเท่ากับ 0.5 และความน่าจะเป็นที่ผู้ป่วยแต่ละคนจะไม่หายจากโรคมะเร็งเป็น $1-0.5=0.5$

จะเห็นว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินาม

1) ความน่าจะเป็นที่จะมีผู้ป่วยหายจากโรคมะเร็งอย่างน้อย 3 คน คือ

$$
\begin{aligned}
P(X \geq 3) &= P(X=3)+P(X=4)+P(X=5)+P(X=6) \\
&= \binom{6}{3}(0.5)^3(0.5)^3+\binom{6}{4}(0.5)^4(0.5)^2 \\
&\quad +\binom{6}{5}(0.5)^5(0.5)+\binom{6}{6}(0.5)^6 \\
&\approx 0.3125+0.2344+0.0938+0.0156 \\
&\approx 0.6563
\end{aligned}
$$

2) เนื่องจาก $\mu_X = 6(0.5) = 3$
ดังนั้น ค่าคาดหมายของจำนวนผู้ป่วยที่หายจากโรคมะเร็ง คือ 3 คน
ซึ่งหมายความว่า ในการสุ่มผู้ป่วยโรคมะเร็งที่มารับการรักษาด้วยสมุนไพรนี้จำนวน 6 คน โดยเฉลี่ยแล้วจะมีผู้ป่วยที่หายจากโรคมะเร็ง 3 คน
เนื่องจาก $\sigma_X = \sqrt{6(0.5)(0.5)} \approx 1.22$
ดังนั้น ส่วนเบี่ยงเบนมาตรฐานของจำนวนผู้ป่วยที่หายจากโรคมะเร็งมีค่าประมาณ 1.22 คน ซึ่งหมายความว่า ในการสุ่มผู้ป่วยโรคมะเร็งที่มารับการรักษาด้วยสมุนไพรนี้จำนวน 6 คน จำนวนผู้ป่วยที่หายจากโรคมะเร็งจะต่างจากค่าคาดหมายประมาณ 1.22 คน ■

ตัวอย่างที่ 12

จากข้อมูลเกี่ยวกับคุณภาพของสินค้าซึ่งเก็บรวบรวมมาในอดีตทำให้ทราบว่า ความน่าจะเป็นที่สินค้าแต่ละชิ้นจะชำรุดเป็น 0.05 และในกระบวนการตรวจสอบคุณภาพสินค้าของโรงงานมีหลักการคือพนักงานจะสุ่มสินค้าจำนวน 5 ชิ้น จากแต่ละกล่องเพื่อตรวจสอบคุณภาพ ถ้าตรวจพบสินค้าชำรุดไม่เกิน 1 ชิ้น สินค้ากล่องนั้นจะผ่านการตรวจสอบคุณภาพ

1) จงหาความน่าจะเป็นที่สินค้าแต่ละกล่องที่ส่งมาตรวจสอบจะผ่านการตรวจสอบคุณภาพ
2) ในการผลิตสินค้าครั้งหนึ่ง ฝ่ายผลิตของโรงงานส่งสินค้ามาให้พนักงานตรวจสอบคุณภาพจำนวน 100 กล่อง จะมีสินค้าที่ผ่านการตรวจสอบคุณภาพกี่กล่อง

วิธีทำ ให้ตัวแปรสุ่ม X คือจำนวนสินค้าที่ชำรุด เมื่อสุ่มสินค้า 5 ชิ้น จากแต่ละกล่อง
จะได้ ค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2, 3, 4 และ 5
เนื่องจากตัวแปรสุ่ม X มีลักษณะดังต่อไปนี้

1. เกิดจากการสุ่มสินค้าจากแต่ละกล่องจำนวน 5 ครั้ง ที่เป็นอิสระกัน
2. การทดลองสุ่มแต่ละครั้งเกิดผลลัพธ์ได้เพียง 2 แบบ คือ สำเร็จ (สินค้าที่สุ่มมาชำรุด) หรือไม่สำเร็จ (สินค้าที่สุ่มมาไม่ชำรุด)
3. ความน่าจะเป็นที่สินค้าชำรุดเมื่อสุ่มสินค้าแต่ละครั้งเท่ากัน โดยเท่ากับ 0.05 และความน่าจะเป็นที่สินค้าไม่ชำรุดเมื่อสุ่มสินค้าแต่ละครั้งเป็น $1 - 0.05 = 0.95$

จะเห็นว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินาม

1) ความน่าจะเป็นที่สินค้าแต่ละกล่องที่ส่งมาตรวจสอบจะผ่านการตรวจสอบคุณภาพ คือ

$$\begin{aligned} P(X \le 1) &= P(X=0)+P(X=1) \\ &= \binom{5}{0}(0.95)^5 + \binom{5}{1}(0.05)(0.95)^4 \\ &\approx 0.7738 + 0.2036 \\ &\approx 0.9774 \end{aligned}$$

2) จากข้อ 1) ความน่าจะเป็นที่สินค้าแต่ละกล่องที่ส่งมาตรวจสอบจะผ่านการตรวจสอบคุณภาพมีค่าประมาณ 0.9774

ดังนั้น ถ้าพนักงานตรวจสอบสินค้าทั้งหมด 100 กล่อง จะมีสินค้าที่ผ่านการตรวจสอบคุณภาพประมาณ $(0.9774)(100) = 97.74$ หรือประมาณ 98 กล่อง

■

แบบฝึกหัด 4.2.2

1. กำหนดให้ $X \sim B(6, 0.3)$ จงหา
   1) $P(X=2)$
   2) $P(X \le 2)$
   3) $P(X > 2)$
   4) $P(2 \le X \le 5)$

2. ในการโยนเหรียญที่ไม่เที่ยงตรงเหรียญหนึ่ง พบว่า ความน่าจะเป็นที่เหรียญขึ้นก้อยในการโยนเหรียญแต่ละครั้งเท่ากับ 0.6 ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่เหรียญขึ้นหัว จากการโยนเหรียญนี้ 6 ครั้ง
   1) จงหาค่าที่เป็นไปได้ของตัวแปรสุ่ม X
   2) จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินามหรือไม่ พร้อมทั้งให้เหตุผลประกอบ
   3) จงหาความน่าจะเป็นที่เหรียญขึ้นหัวน้อยกว่า 3 ครั้ง

4) โดยเฉลี่ยแล้วเหรียญจะขึ้นหัวกี่ครั้ง
5) จงหาความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

3. ให้ตัวแปรสุ่ม Y คือจำนวนครั้งที่ได้แต้มเป็นจำนวนคู่ จากการทอดลูกเต๋าที่เที่ยงตรง 1 ลูก 8 ครั้ง
   1) จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y เป็นการแจกแจงทวินามหรือไม่ พร้อมทั้งให้เหตุผลประกอบ
   2) จงหาความน่าจะเป็นที่ได้แต้มเป็นจำนวนคู่ 5 ครั้ง
   3) จงหาความน่าจะเป็นที่ได้แต้มเป็นจำนวนคู่น้อยกว่า 8 ครั้ง
   4) จงหาค่าคาดหมายและความแปรปรวนของตัวแปรสุ่ม Y

4. ความน่าจะเป็นที่โสภิตาจะซื้อชานมไข่มุกในแต่ละวันเท่ากับ $\frac{9}{10}$ จงหาความน่าจะเป็นที่โสภิตาจะซื้อชานมไข่มุกไม่เกิน 2 วัน ในหนึ่งสัปดาห์

5. ในการแข่งขันตอบโจทย์ปัญหาทางวิชาการของโรงเรียนแห่งหนึ่งมีผู้เข้าร่วมการแข่งขันจำนวน 6 คน ทำการแข่งขันทั้งหมด 5 ครั้ง ถ้าภัคนินทร์เป็นหนึ่งในผู้เข้าแข่งขันและความน่าจะเป็นที่ภัคนินทร์จะชนะการแข่งขันในแต่ละครั้งเท่ากันโดยเท่ากับ 0.3 จงหาความน่าจะเป็นที่ภัคนินทร์จะชนะการแข่งขันอย่างน้อย 1 ครั้ง

6. จากข้อมูลเกี่ยวกับปัญหาที่เกิดขึ้นในกระบวนการผลิตแผงวงจรไฟฟ้าของโรงงานผลิตอุปกรณ์อิเล็กทรอนิกส์แห่งหนึ่ง พบว่า แผงวงจรไฟฟ้าที่ชำรุดแต่ละแผ่นเกิดจากสาเหตุใดสาเหตุหนึ่งด้วยความน่าจะเป็น ดังต่อไปนี้

สาเหตุ	ความน่าจะเป็น
โลหะบัดกรีเป็นรู	0.4
อุปกรณ์เชื่อมไม่ติด	0.3
อุปกรณ์แตกร้าว	0.2
อื่น ๆ	0.1

ถ้าพนักงานสุ่มแผงวงจรไฟฟ้าที่ชำรุดจำนวน 3 แผ่น จงหา

1) ความน่าจะเป็นที่แผงวงจรไฟฟ้าทั้งสามแผ่นชำรุดเนื่องจากโลหะบัดกรีเป็นรู
2) ความน่าจะเป็นที่มีแผงวงจรไฟฟ้า 2 แผ่น ชำรุดเนื่องจากอุปกรณ์เชื่อมไม่ติด
3) ค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของจำนวนแผงวงจรไฟฟ้าที่ชำรุดเนื่องจากอุปกรณ์แตกร้าว

7. จากข้อมูลของศูนย์ควบคุมและสั่งการจราจร พบว่า ความน่าจะเป็นที่รถยนต์แต่ละคันจะเปลี่ยนช่องทางเดินรถในพื้นที่ห้าม (เส้นทึบ) บริเวณสี่แยกไฟแดงแห่งหนึ่งเป็น 0.75 ถ้าสุ่มรถยนต์ที่วิ่งผ่านป้อมควบคุมสัญญาณไฟจราจรบริเวณสี่แยกนี้มา 9 คัน จงหา

1) ความน่าจะเป็นที่จะพบรถยนต์เปลี่ยนช่องทางเดินรถในพื้นที่ห้าม (เส้นทึบ) 4 คัน
2) ความน่าจะเป็นที่จะพบรถยนต์เปลี่ยนช่องทางเดินรถในพื้นที่ห้าม (เส้นทึบ) ไม่เกิน 3 คัน
3) ความน่าจะเป็นที่จะพบรถยนต์เปลี่ยนช่องทางเดินรถในพื้นที่ห้าม (เส้นทึบ) มากกว่า 6 คัน
4) ค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของจำนวนรถยนต์ที่เปลี่ยนช่องทางเดินรถในพื้นที่ห้าม (เส้นทึบ)

8. สาเหตุหนึ่งของภาวะคอเลสเตอรอลสูงเกิดจากมิวเทชันของยีน *LDLR* (low-density lipoprotein receptor) ที่เกี่ยวข้องกับการสร้างตัวรับ LDL ที่บริเวณเยื่อหุ้มเซลล์ ซึ่งส่งผลต่อระดับคอเลสเตอรอลในเลือด ดังรายละเอียดต่อไปนี้

บุคคลที่มีจีโนไทป์ L^HL^H สามารถสร้างตัวรับ LDL ได้

บุคคลที่มีจีโนไทป์ L^HL^h สามารถสร้างตัวรับ LDL ได้ในปริมาณน้อย ส่งผลให้มีโอกาสมีระดับคอเลสเตอรอลในเลือดค่อนข้างสูง

บุคคลที่มีจีโนไทป์ L^hL^h **ไม่**สามารถสร้างตัวรับ LDL ได้ ส่งผลให้มีระดับคอเลสเตอรอลในเลือดสูงมาก และมีโอกาสเป็นโรคหัวใจตั้งแต่อายุยังน้อยได้

สำหรับพ่อและแม่ที่มีจีโนไทป์ L^HL^h ความน่าจะเป็นที่ลูกแต่ละคนจะมีจีโนไทป์ L^HL^H และ L^HL^h คือ $\frac{1}{4}$ และ $\frac{1}{2}$ ตามลำดับ ถ้าสามีภรรยาคู่หนึ่งที่มีจีโนไทป์ L^HL^h ทั้งคู่ ต้องการมีบุตร 3 คน จงหา

1) ความน่าจะเป็นที่บุตรทั้ง 3 คน ไม่มีจีโนไทป์ L^hL^h
2) ความน่าจะเป็นที่มีบุตรอย่างน้อย 1 คน มีจีโนไทป์ L^hL^h

4.3 การแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง

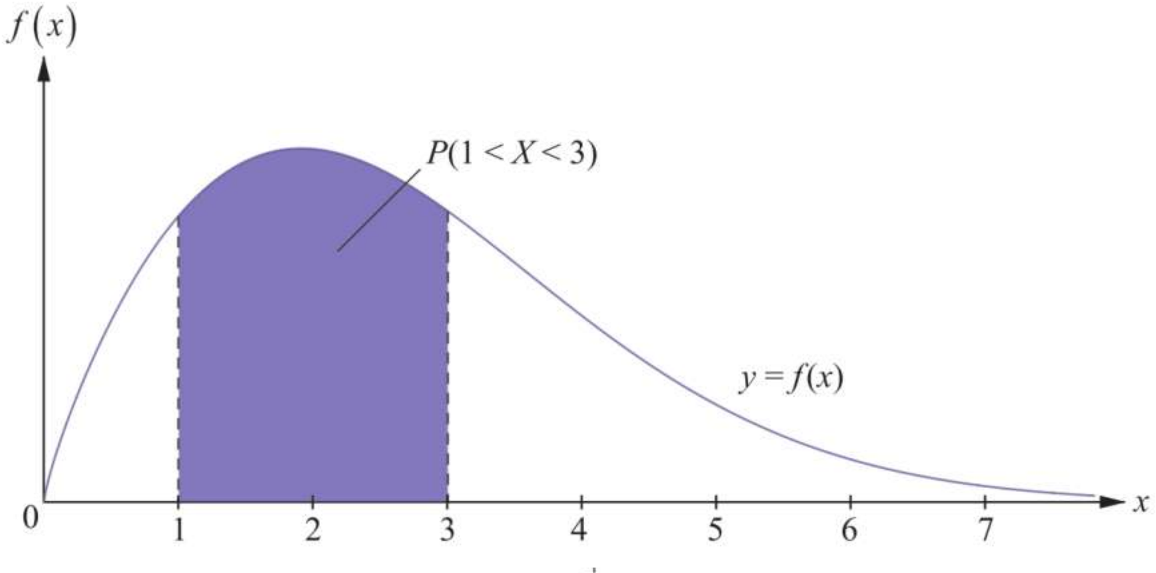
ในกรณีที่ตัวแปรสุ่มที่สนใจเป็นตัวแปรสุ่มต่อเนื่อง เช่น ระยะเวลาที่ลูกค้าใช้บริการในห้างสรรพสินค้าแห่งหนึ่ง น้ำหนักของผู้ป่วยที่เข้ารับการรักษาที่โรงพยาบาลแห่งหนึ่ง เนื่องจากตัวแปรสุ่มต่อเนื่องมีเซตของค่าที่เป็นไปได้ทั้งหมดเป็นช่วงซึ่งเป็นสับเซตของ $\mathbb{R}$ ซึ่งมีสมาชิกเป็นจำนวนอนันต์ จึงไม่เหมาะกับการเขียนแสดงการแจกแจงความน่าจะเป็นในรูปตาราง แต่จะใช้ **เส้นโค้งความหนาแน่น (density curve)** ในการเขียนแสดงการแจกแจงความน่าจะเป็น โดยความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงใดช่วงหนึ่งจะเท่ากับพื้นที่ที่ปิดล้อมด้วยเส้นโค้งความหนาแน่นกับแกน X ในช่วงนั้น จะเรียกพื้นที่บริเวณดังกล่าวว่า **พื้นที่ใต้เส้นโค้งความหนาแน่น**

เส้นโค้งความหนาแน่นเป็นกราฟของฟังก์ชัน $y = f(x)$ โดยที่ x แทนค่าที่เป็นไปได้ของตัวแปรสุ่ม จะเรียกฟังก์ชันนี้ว่า **ฟังก์ชันความหนาแน่นความน่าจะเป็น (probability density function)**

หมายเหตุ $f(x)$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่ม X ก็ต่อเมื่อ

1. $f(x) \geq 0$ สำหรับทุก x ที่เป็นค่าที่เป็นไปได้ของตัวแปรสุ่ม X
2. พื้นที่ใต้เส้นโค้งความหนาแน่นทั้งหมดจะเท่ากับ 1

พิจารณาเส้นโค้งความหนาแน่นของตัวแปรสุ่ม X ดังรูป



รูปที่ 1

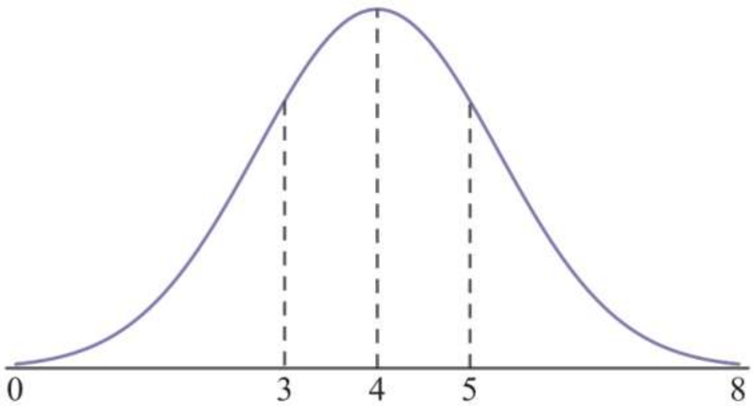
จากรูป สามารถหาความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงที่สนใจได้จากการหาพื้นที่ใต้เส้นโค้ง $y=f(x)$ ในช่วงดังกล่าว เช่น ความน่าจะเป็นที่ตัวแปรสุ่ม X มีค่ามากกว่า 1 แต่น้อยกว่า 3 จะเท่ากับพื้นที่ส่วนที่แรเงา นั่นคือ $P(1<X<3)=\int_1^3 f(x)\,dx$

ดังนั้น ถ้าทราบฟังก์ชันความหนาแน่นความน่าจะเป็น แล้วจะสามารถหาความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงใดช่วงหนึ่งได้โดยการหาปริพันธ์จำกัดเขตของฟังก์ชันในช่วงดังกล่าว

ถ้าให้ X เป็นตัวแปรสุ่มต่อเนื่อง และ a เป็นค่าที่เป็นไปได้ของ X จะได้ว่า $P(X=a)=0$ เนื่องจากพื้นที่ใต้เส้นโค้งความหนาแน่นจาก a ถึง a เท่ากับศูนย์ ดังนั้น สำหรับตัวแปรสุ่มต่อเนื่อง จะไม่พิจารณาความน่าจะเป็นของการเกิดค่าของตัวแปรสุ่มค่าใดค่าหนึ่ง แต่จะสนใจเฉพาะความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงใดช่วงหนึ่ง โดยความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงปิด $[a, b]$ จะเท่ากับความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงเปิด (a, b) นั่นคือ $P(a \le X \le b)=P(a<X<b)$ เมื่อ a และ b เป็นค่าที่เป็นไปได้ของตัวแปรสุ่ม X ในทำนองเดียวกัน จะได้ว่า $P(X \le a)=P(X<a)$ และ $P(X \ge a)=P(X>a)$ เมื่อ a เป็นค่าที่เป็นไปได้ของตัวแปรสุ่ม X

4.3.1 การแจกแจงปกติ

เส้นโค้งความหนาแน่นที่พบบ่อยมักมีลักษณะสมมาตรคล้ายรูประฆัง เช่น ถ้าให้ตัวแปรสุ่มคือระยะเวลาที่นักเรียนใช้ในห้องสมุดในแต่ละวัน ซึ่งอาจมีค่าเป็นค่าใดก็ได้ในช่วง 0 ถึง 8 ชั่วโมง จะได้ว่าตัวแปรสุ่มนี้เป็นตัวแปรสุ่มต่อเนื่อง นอกจากนี้มักพบว่าจำนวนนักเรียนที่ใช้เวลาในห้องสมุดน้อยกว่า 1 ชั่วโมง มีน้อยมาก และจำนวนนักเรียนที่ใช้เวลาในห้องสมุดมากกว่า 7 ชั่วโมง ก็มีน้อยมาก ส่วนใหญ่แล้วนักเรียนจะใช้เวลาในห้องสมุดประมาณ $3-5$ ชั่วโมง ดังนั้น เมื่อเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มด้วยเส้นโค้งความหนาแน่น จะได้เส้นโค้งที่โด่งกลางแล้วลาดลงทั้งสองด้าน ดังรูป



การแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีลักษณะเช่นนี้เรียกว่า **การแจกแจงปกติ** ซึ่งมีบทนิยามดังนี้

บทนิยาม 5

การแจกแจงปกติ (normal distribution) คือ การแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง X ที่มีฟังก์ชันความหนาแน่นความน่าจะเป็น คือ

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{เมื่อ} \quad -\infty < x < \infty$$

โดยที่ μ แทนค่าเฉลี่ย

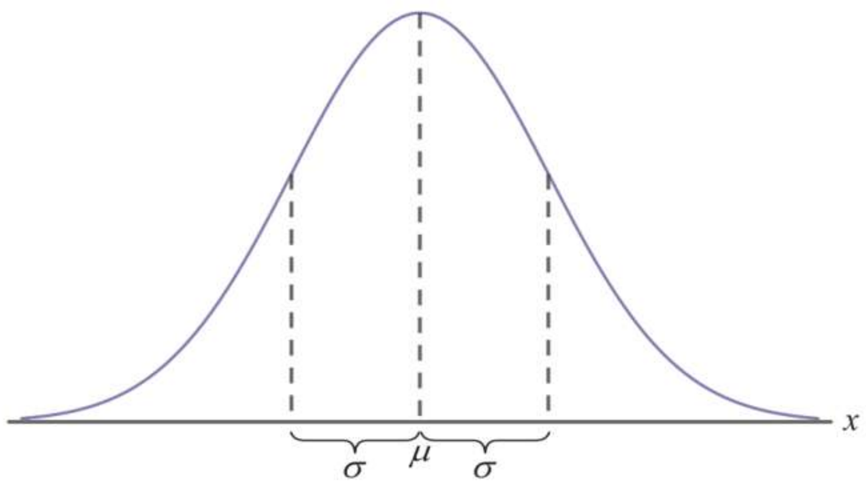
และ σ แทนส่วนเบี่ยงเบนมาตรฐาน

การแจกแจงปกติเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่องที่มีความสำคัญและใช้มากในสถิติศาสตร์ เนื่องจากเป็นการแจกแจงที่มีความใกล้เคียงกับข้อมูลที่เกิดตามธรรมชาติหรือที่มนุษย์สร้างขึ้น เช่น ความสูงของประชากรไทย คะแนนสอบ O-NET วิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2562 อายุการใช้งานของสินค้าที่ผลิตจากโรงงานแห่งหนึ่ง

ถ้าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงปกติ แล้วเมื่อเขียนกราฟของฟังก์ชันความหนาแน่นความน่าจะเป็นสำหรับตัวแปรสุ่ม X จะได้ **เส้นโค้งปกติ (normal curve)** ซึ่งเป็นเส้นโค้งรูปประฆังที่มีสมบัติดังต่อไปนี้

1. เส้นโค้งมีเส้นตั้งฉากกับแกน X ที่ลากผ่านค่าเฉลี่ยเป็นแกนสมมาตร ทำให้พื้นที่ใต้เส้นโค้งทางด้านซ้ายของค่าเฉลี่ยเท่ากับพื้นที่ใต้เส้นโค้งทางด้านขวาของค่าเฉลี่ย
2. ปลายเส้นโค้งทั้งสองด้านเข้าใกล้แกน X แต่จะไม่ตัดแกน X หรือกล่าวได้ว่าแกน X เป็นเส้นกำกับแนวนอน
3. ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (หรือความแปรปรวน) จะเป็นตัวกำหนดลักษณะเฉพาะของเส้นโค้งว่ามีแกนสมมาตรอยู่ที่ใด และมีการกระจายจากค่าเฉลี่ยมากน้อยเพียงใด

ตัวอย่างเช่น เส้นโค้งปกติที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ สามารถเขียนได้ดังรูป

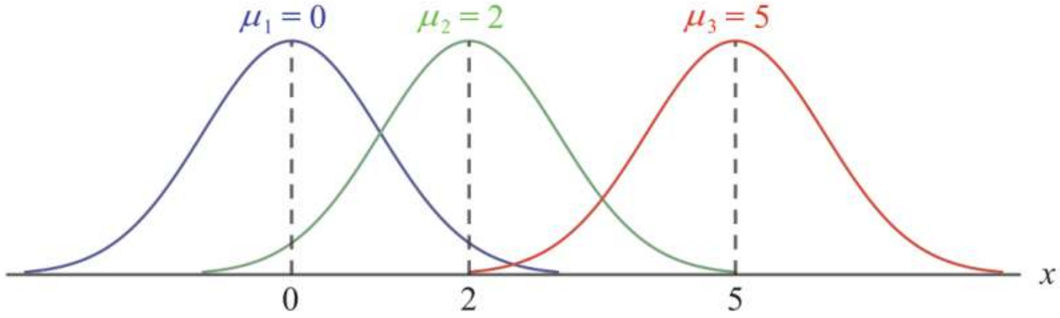


รูปที่ 2

ถ้าค่าเฉลี่ยหรือส่วนเบี่ยงเบนมาตรฐานค่าใดค่าหนึ่งหรือทั้งสองค่าเปลี่ยนแปลงไป เส้นโค้งปกติจะเปลี่ยนแปลงตามไปด้วย แต่ยังคงเป็นเส้นโค้งรูปประฆัง ดังตัวอย่างต่อไปนี้

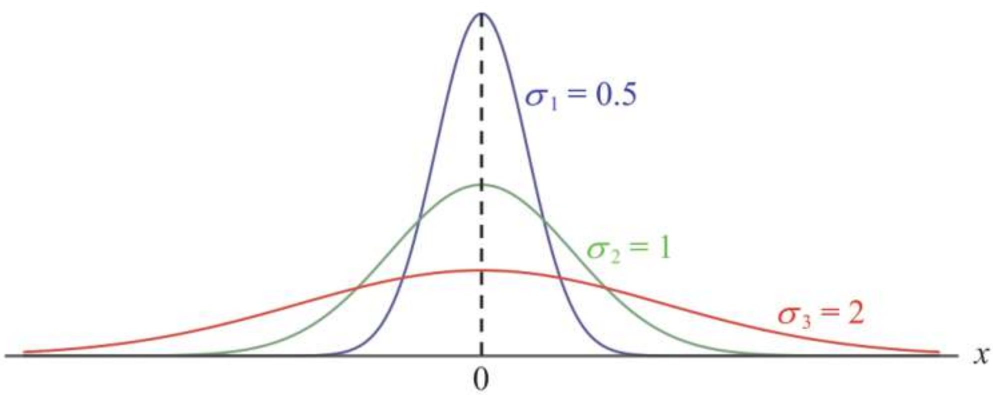
เมื่อค่าเฉลี่ยแตกต่างกัน แต่ส่วนเบี่ยงเบนมาตรฐานเท่ากัน

$\sigma = 1$



เมื่อส่วนเบี่ยงเบนมาตรฐานแตกต่างกัน แต่ค่าเฉลี่ยเท่ากัน

$\mu = 0$



ถ้าตัวแปรสุ่ม X มีการแจกแจงปกติ โดยที่ μ แทนค่าเฉลี่ย และ σ^2 แทนความแปรปรวน จะเรียกตัวแปรสุ่ม X ว่า **ตัวแปรสุ่มปกติ** เรียก μ และ σ^2 ว่า **พารามิเตอร์ของการแจกแจงปกติ** และเขียนสัญลักษณ์ $X \sim N(\mu, \sigma^2)$ เพื่อแสดงว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงปกติที่มี μ และ σ^2 เป็นพารามิเตอร์

จากที่กล่าวมาแล้วว่าสามารถหาความน่าจะเป็นที่ตัวแปรสุ่มจะมีค่าอยู่ในช่วงที่สนใจได้จากการหาพื้นที่ใต้เส้นโค้งความหนาแน่นในช่วงนั้น ซึ่งเท่ากับปริพันธ์จำกัดเขตของฟังก์ชันความหนาแน่นความน่าจะเป็นในช่วงดังกล่าว โดยจะต้องใช้วิธีการของแคลคูลัสซึ่งค่อนข้างยุ่งยาก ในทางปฏิบัติจึงหาพื้นที่ใต้เส้นโค้งปกติโดยใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติ แต่เนื่องจากเป็นไปไม่ได้ที่จะสร้างตารางหลาย ๆ ตารางมาแสดงพื้นที่ใต้เส้นโค้งปกติซึ่งมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานต่างกัน ดังนั้นจะใช้วิธีการแปลงการแจกแจงปกติของตัวแปรสุ่มให้เป็นการแจกแจงปกติมาตรฐาน ซึ่งจะกล่าวถึงในหัวข้อต่อไป

4.3.2 การแจกแจงปกติมาตรฐาน

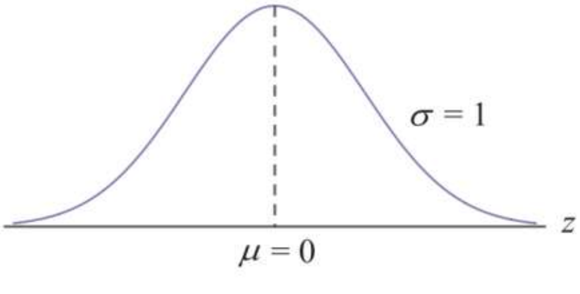
บทนิยาม 6

การแจกแจงปกติมาตรฐาน (standard normal distribution) คือ การแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 $(\mu = 0)$ และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 $(\sigma = 1)$

จะได้ว่าฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่ม Z ที่มีการแจกแจงปกติมาตรฐาน คือ

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{เมื่อ} \quad -\infty < z < \infty$$

เรียกเส้นโค้งปกติซึ่งได้จากตัวแปรสุ่มปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ว่า **เส้นโค้งปกติมาตรฐาน (standard normal curve)** ดังรูป



เรียกตัวแปรสุ่มที่มีการแจกแจงปกติมาตรฐานว่า **ตัวแปรสุ่มปกติมาตรฐาน (standard normal random variable)**

สำหรับการหาความน่าจะเป็นที่ตัวแปรสุ่มปกติมาตรฐานจะมีค่าอยู่ในช่วงที่สนใจ จะใช้ตารางแสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน (ตารางที่ 1) แทนการหาปริพันธ์จำกัดเขตของฟังก์ชันความหนาแน่นความน่าจะเป็น โดยค่าที่ปรากฏในตารางที่ 1 คือค่าประมาณของพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง z หรือความน่าจะเป็นที่ตัวแปรสุ่มปกติมาตรฐาน Z มีค่าน้อยกว่าหรือเท่ากับ z เขียนแทนด้วยสัญลักษณ์ $P(Z \le z)$

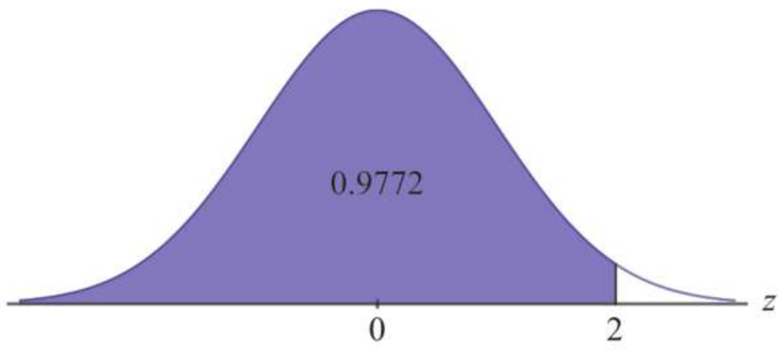
การอ่านตารางที่ 1 ให้เริ่มพิจารณาจากแถวที่แสดงค่า z จาก 0.0 ถึง –3.0 หรือจาก 0.0 ถึง 3.0 ค่าของ z มีค่าลดลงหรือเพิ่มขึ้นแถวละ 0.1 จากนั้นจึงพิจารณาหลักซึ่งแสดงทศนิยมตำแหน่งที่ 2 ของค่า z เช่น ถ้าต้องการหา $P(Z \le -1.54)$ จะเริ่มพิจารณาจากแถวที่แสดงค่า $z = -1.5$ จากนั้นพิจารณาหลักที่แสดงค่า 0.04 จะได้ว่าพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง -1.54 หรือ $P(Z \le -1.54)$ มีค่าประมาณ 0.0618

ตัวอย่างที่ 13

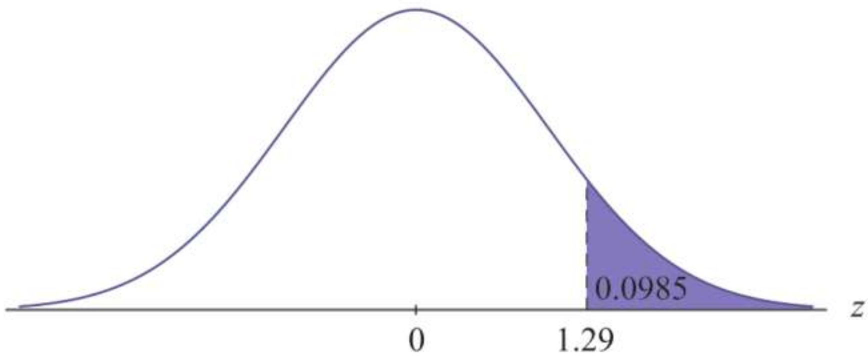
ให้ Z เป็นตัวแปรสุ่มปกติมาตรฐาน จงหา

1) $P(Z \le 2)$
2) $P(Z > 1.29)$
3) $P(-1.27 \le Z \le 0.45)$

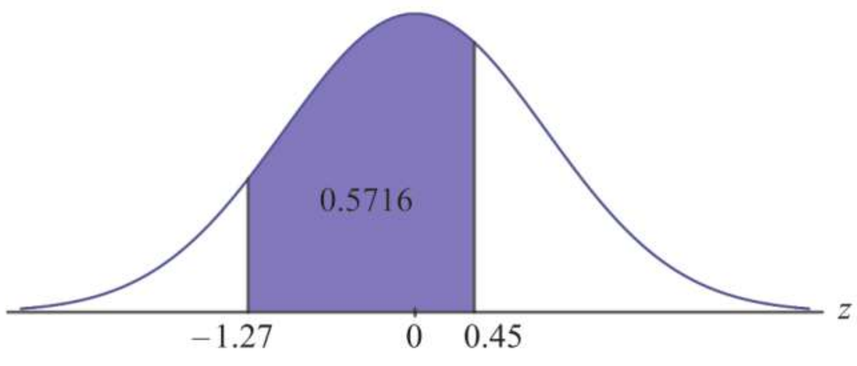
วิธีทำ 1) จากตารางที่ 1 อ่านค่าประมาณของพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง 2 ได้เป็น 0.9772 นั่นคือ $P(Z \le 2) = 0.9772$



2) เนื่องจาก $P(Z > 1.29) = 1 - P(Z \le 1.29)$
และจากตารางที่ 1 อ่านค่าประมาณของพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง 1.29 ได้เป็น 0.9015 นั่นคือ $P(Z \le 1.29) = 0.9015$
ดังนั้น $P(Z > 1.29) = 1 - 0.9015 = 0.0985$



3) เนื่องจาก $P(-1.27 \le Z \le 0.45) = P(Z \le 0.45) - P(Z < -1.27)$
และจากตารางที่ 1 อ่านค่าประมาณของพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง 0.45 ได้เป็น 0.6736 นั่นคือ $P(Z \le 0.45) = 0.6736$
และอ่านค่าประมาณของพื้นที่ใต้เส้นโค้งปกติมาตรฐานจาก $-\infty$ ถึง -1.27 ได้เป็น 0.1020 นั่นคือ $P(Z < -1.27) = P(Z \le -1.27) = 0.1020$
ดังนั้น $P(-1.27 \le Z \le 0.45) = 0.6736 - 0.1020 = 0.5716$



■

ในกรณีที่ตัวแปรสุ่มมีการแจกแจงปกติแต่ไม่ใช่การแจกแจงปกติมาตรฐาน จะไม่สามารถใช้ตารางที่ 1 ในการหาความน่าจะเป็นได้ ดังนั้นจะต้องแปลงตัวแปรสุ่มปกติให้เป็นตัวแปรสุ่มปกติมาตรฐาน โดยใช้ทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2

ให้ตัวแปรสุ่ม X มีการแจกแจงปกติ โดยมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ ถ้าตัวแปรสุ่ม Z นิยามโดย $Z = \dfrac{X-\mu}{\sigma}$ แล้ว ตัวแปรสุ่ม Z จะมีการแจกแจงปกติมาตรฐาน นั่นคือ $\mu_Z = 0$ และ $\sigma_Z = 1$

นอกจากนี้ $P(a \le X \le b) = P\left(\dfrac{a-\mu}{\sigma} \le Z \le \dfrac{b-\mu}{\sigma}\right)$ เมื่อ a, b เป็นค่าที่เป็นไปได้ของตัวแปรสุ่ม X และ $a \le b$

ตัวอย่างที่ 14

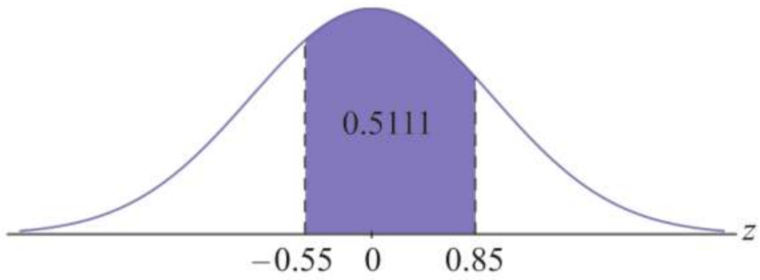
กำหนดให้ $X \sim N(3.5, 4)$ จงหา $P(2.4 < X < 5.2)$

วิธีทำ เนื่องจาก $X \sim N(3.5, 4)$ จะได้ว่าตัวแปรสุ่ม X มีการแจกแจงปกติ โดยที่ $\mu = 3.5$ และ $\sigma^2 = 4$ นั่นคือ $\sigma = 2$

ให้ $Z = \dfrac{X-\mu}{\sigma}$

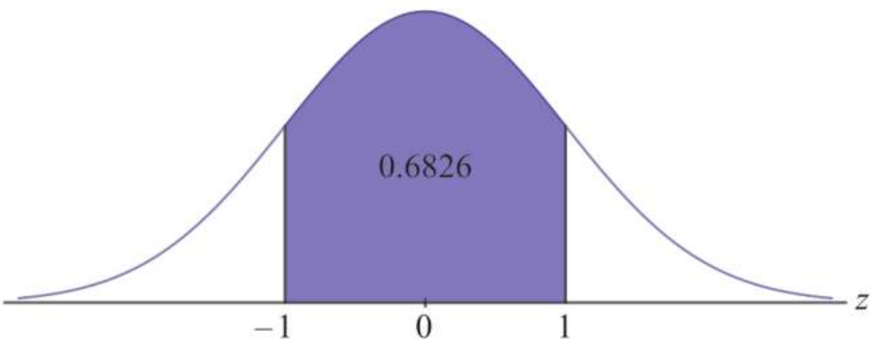
ดังนั้น

$$\begin{aligned} P(2.4 < X < 5.2) &= P\left(\frac{2.4-3.5}{2} < \frac{X-\mu}{\sigma} < \frac{5.2-3.5}{2}\right) \\ &= P(-0.55 < Z < 0.85) \\ &= P(Z < 0.85) - P(Z \le -0.55) \\ &= 0.8023 - 0.2912 \\ &= 0.5111 \end{aligned}$$

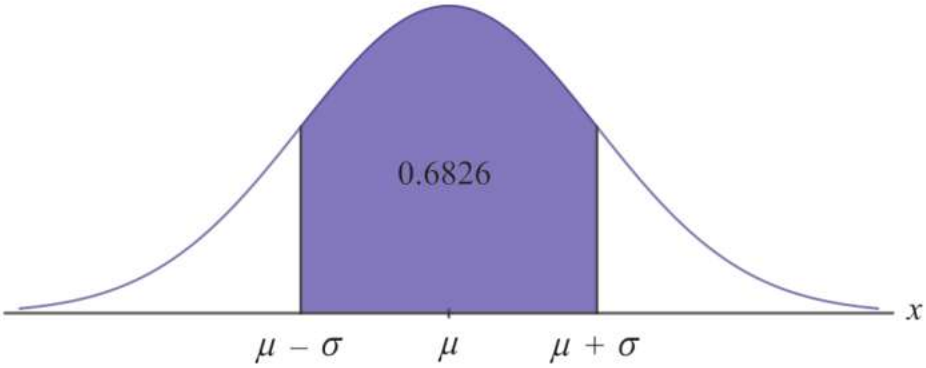


สำหรับตัวแปรสุ่มปกติ X ที่มีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ และตัวแปรสุ่ม Z นิยามโดย $Z = \dfrac{X-\mu}{\sigma}$ จะได้ว่า

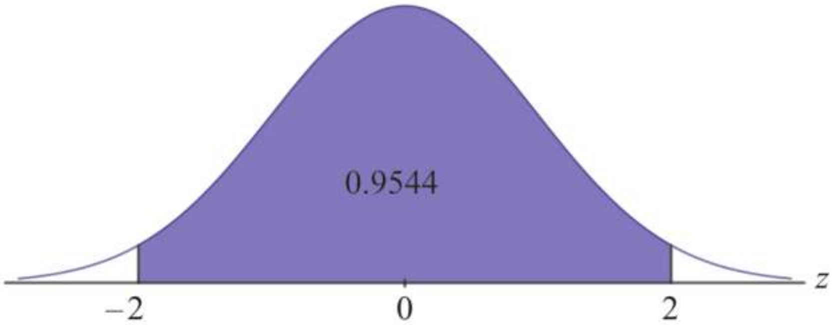
1. $$\begin{aligned} P(\mu-\sigma \le X \le \mu+\sigma) &= P(-1 \le Z \le 1) \\ &= P(Z \le 1) - P(Z < -1) \\ &= 0.8413 - 0.1587 \\ &= 0.6826 \end{aligned}$$



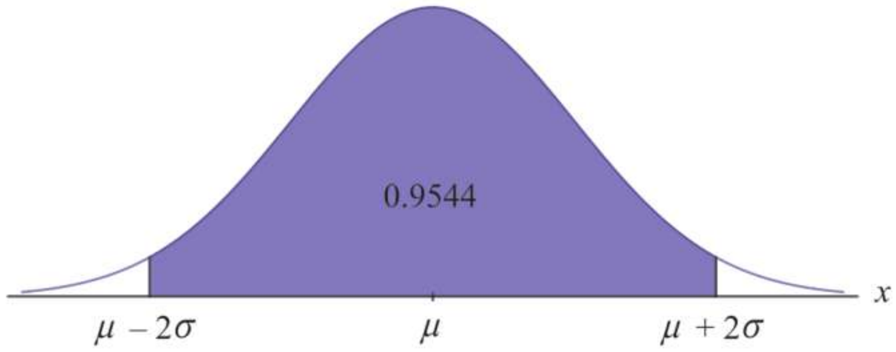
นั่นคือ ความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าอยู่ในช่วง $[\mu-\sigma, \mu+\sigma]$ มีค่าประมาณ 0.6826 หรือพื้นที่ใต้เส้นโค้งปกติจาก $\mu-\sigma$ ถึง $\mu+\sigma$ มีค่าประมาณ 68.26% ของพื้นที่ใต้เส้นโค้งปกติทั้งหมด



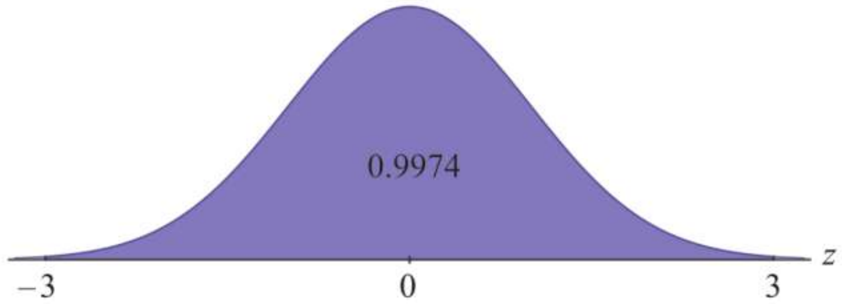
2. $$\begin{aligned} P(\mu-2\sigma \le X \le \mu+2\sigma) &= P(-2 \le Z \le 2) \\ &= P(Z \le 2) - P(Z < -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$



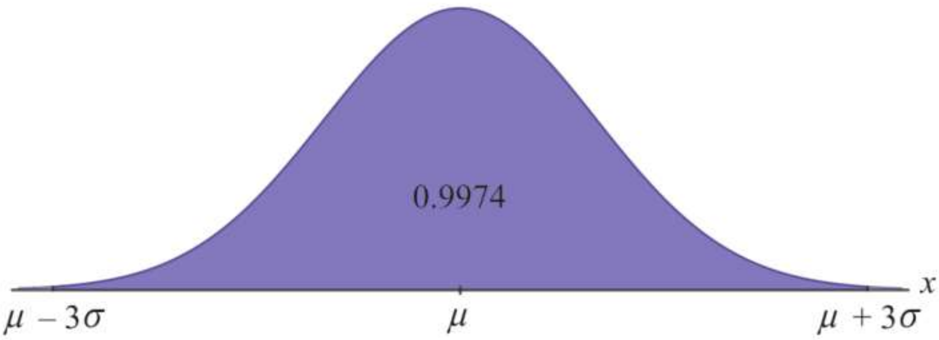
นั่นคือ ความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าอยู่ในช่วง $[\mu-2\sigma, \mu+2\sigma]$ มีค่าประมาณ 0.9544 หรือพื้นที่ใต้เส้นโค้งปกติจาก $\mu-2\sigma$ ถึง $\mu+2\sigma$ มีค่าประมาณ 95.44% ของพื้นที่ใต้เส้นโค้งปกติทั้งหมด



3. $$\begin{aligned} P(\mu-3\sigma \le X \le \mu+3\sigma) &= P(-3 \le Z \le 3) \\ &= P(Z \le 3) - P(Z < -3) \\ &= 0.9987 - 0.0013 \\ &= 0.9974 \end{aligned}$$



นั่นคือ ความน่าจะเป็นที่ตัวแปรสุ่ม X จะมีค่าอยู่ในช่วง $[\mu-3\sigma, \mu+3\sigma]$ มีค่าประมาณ 0.9974 หรือพื้นที่ใต้เส้นโค้งปกติจาก $\mu-3\sigma$ ถึง $\mu+3\sigma$ มีค่าประมาณ 99.74% ของพื้นที่ใต้เส้นโค้งปกติทั้งหมด



ตัวอย่างที่ 15

ความสูงของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 160 และ 5 เซนติเมตร ตามลำดับ ถ้าสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำนวน 1 คนจากโรงเรียนนี้ จงหาความน่าจะเป็นที่นักเรียนที่สุ่มได้จะมีความสูง

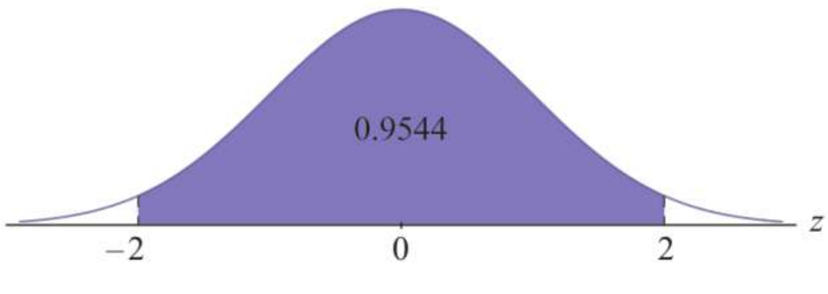
1) ระหว่าง 150 และ 170 เซนติเมตร
2) มากกว่า 162 เซนติเมตร

วิธีทำ ให้ตัวแปรสุ่ม X คือความสูงของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งนี้ จะได้ว่าตัวแปรสุ่ม X มีการแจกแจงปกติ โดยที่ $\mu = 160$ และ $\sigma = 5$

ให้ $Z = \dfrac{X - \mu}{\sigma}$

1) เนื่องจาก

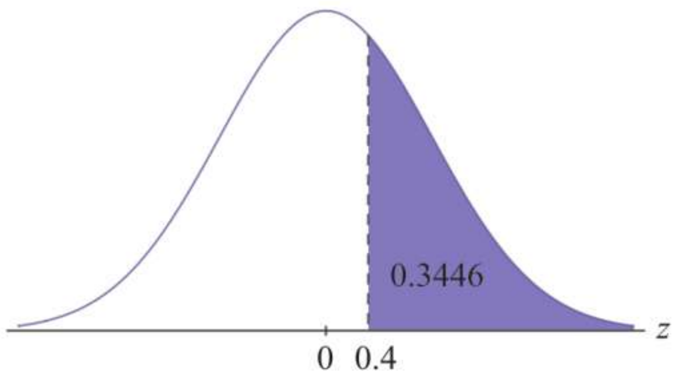
$$\begin{aligned} P(150 < X < 170) &= P\left(\frac{150-160}{5} < \frac{X-\mu}{\sigma} < \frac{170-160}{5}\right) \\ &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z \le -2) \\ &= 0.9772 - 0.0228 \\ &= 0.9544 \end{aligned}$$



ดังนั้น ความน่าจะเป็นที่นักเรียนที่สุ่มได้จะมีความสูงระหว่าง 150 และ 170 เซนติเมตร คือ 0.9544

2) เนื่องจาก

$$\begin{aligned} P(X>162) &= P\left(Z>\frac{162-160}{5}\right) \\ &= P(Z>0.4) \\ &= 1-P(Z\le 0.4) \\ &= 1-0.6554 \\ &= 0.3446 \end{aligned}$$

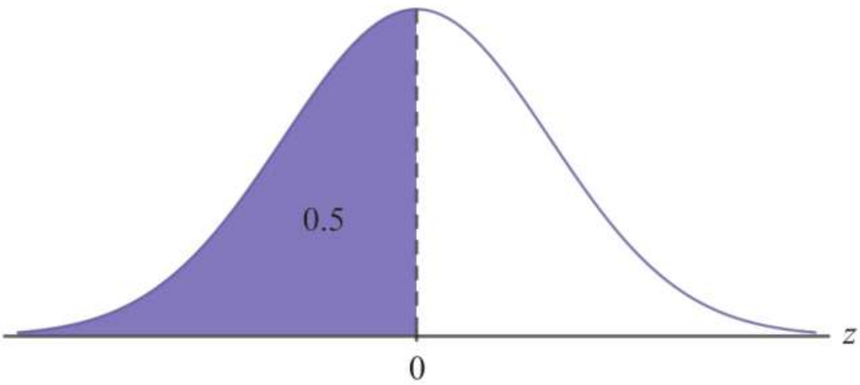


ดังนั้น ความน่าจะเป็นที่นักเรียนที่สุ่มได้จะมีความสูงมากกว่า 162 เซนติเมตร คือ 0.3446 ■

เปอร์เซ็นไทล์ของตัวแปรสุ่มต่อเนื่อง

จากหัวข้อ 3.3 นักเรียนทราบแล้วว่าเปอร์เซ็นไทล์เป็นค่าวัดตำแหน่งที่ของข้อมูลเชิงปริมาณ โดยแบ่งข้อมูลที่เรียงจากน้อยไปมากออกเป็น 100 ส่วน เท่า ๆ กัน สำหรับตัวแปรสุ่มต่อเนื่อง X เนื่องจากพื้นที่ใต้เส้นโค้งความหนาแน่นทั้งหมดเท่ากับ 1 หรือคิดเป็น 100% ดังนั้น ถ้า x เป็นค่าที่เป็นไปได้ของตัวแปรสุ่ม X จะได้ว่าข้อมูลที่มีค่าน้อยกว่า x มีจำนวน $P(X<x)\cdot 100\%$ นั่นคือ ถ้า $P(X<x)\cdot 100$ เป็นจำนวนเต็มที่อยู่ระหว่าง 0 และ 100 จะได้ว่าเปอร์เซ็นไทล์ที่ $P(X<x)\cdot 100$ เท่ากับ x

ตัวอย่างเช่น ถ้า Z เป็นตัวแปรสุ่มปกติมาตรฐาน เนื่องจาก $P(Z<0)=0.5$ ดังนั้น 0 คือเปอร์เซ็นไทล์ที่ $P(Z<0)\cdot 100=(0.5)(100)=50$ หรือกล่าวได้ว่าข้อมูลที่มีค่าน้อยกว่า 0 มีจำนวน 50% ของข้อมูลทั้งหมด โดยเมื่อพิจารณาจากรูปต่อไปนี้จะเห็นว่าบริเวณที่แรเงามีพื้นที่เป็นครึ่งหนึ่งของพื้นที่ใต้เส้นโค้งปกติมาตรฐานทั้งหมด



ตัวอย่างที่ 16

อายุการใช้งานของถ่านไฟฉายชนิดหนึ่งมีการแจกแจงปกติ โดยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 756 และ 35 นาที ตามลำดับ จงหาว่า

1) ถ่านไฟฉายที่มีอายุการใช้งานน้อยกว่า 791 นาที มีกี่เปอร์เซ็นต์ของถ่านไฟฉายทั้งหมด
2) ถ่านไฟฉายที่มีอายุการใช้งานมากกว่าหรือเท่ากับเปอร์เซ็นไทล์ที่ 95 สามารถใช้งานได้อย่างน้อยกี่นาที เมื่อกำหนดให้ $P(Z < 1.645) = 0.95$

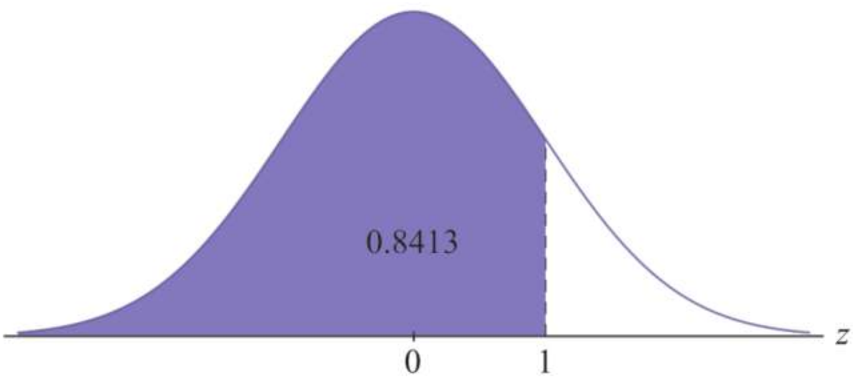
วิธีทำ ให้ตัวแปรสุ่ม X คืออายุการใช้งานของถ่านไฟฉาย

จะได้ว่าตัวแปรสุ่ม X มีการแจกแจงปกติ โดยที่ $\mu = 756$ และ $\sigma = 35$

ให้ $Z = \frac{X-\mu}{\sigma}$

1) เนื่องจาก

$$\begin{aligned} P(X < 791) &= P\left(\frac{X-\mu}{\sigma} < \frac{791-756}{35}\right) \\ &= P(Z < 1) \\ &= 0.8413 \end{aligned}$$



ดังนั้น ถ่านไฟฉายที่มีอายุการใช้งานน้อยกว่า 791 นาที มีประมาณ $(0.8413)(100) = 84.13$ เปอร์เซ็นต์ของถ่ายไฟฉายทั้งหมด

2) เนื่องจากอายุการใช้งานของถ่านไฟฉายที่น้อยที่สุดที่มากกว่าหรือเท่ากับเปอร์เซ็นไทล์ที่ 95 ก็คือเปอร์เซ็นไทล์ที่ 95 ดังนั้น ต้องการหา x ที่ทำให้ $P(X < x) \cdot 100 = 95$

นั่นคือ $P(X < x) = 0.95$

จะได้ $P\left(Z < \dfrac{x - 756}{35}\right) = 0.95$

เนื่องจาก $P(Z < 1.645) = 0.95$

ดังนั้น $\dfrac{x - 756}{35} = 1.645$

$x = 813.575$

นั่นคือ ถ่านไฟฉายที่มีอายุการใช้งานมากกว่าหรือเท่ากับเปอร์เซ็นไทล์ที่ 95 สามารถใช้งานได้อย่างน้อย 813.575 นาที ■

การเปรียบเทียบตำแหน่งของข้อมูลโดยใช้ค่าของตัวแปรสุ่มปกติมาตรฐาน

การแปลงตัวแปรสุ่มปกติให้เป็นตัวแปรสุ่มปกติมาตรฐาน นอกจากจะมีประโยชน์ในการหาความน่าจะเป็นโดยใช้ตารางแล้ว ยังสามารถนำค่าของตัวแปรสุ่มปกติมาตรฐานที่แปลงได้ไปใช้ในการเปรียบเทียบข้อมูลตั้งแต่สองชุดขึ้นไปว่ามีความแตกต่างกันหรือไม่เพียงใด เนื่องจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของข้อมูลแต่ละชุดมักจะไม่เท่ากัน บางครั้งจึงไม่สามารถนำข้อมูลแต่ละชุดมาเปรียบเทียบโดยตรงได้ เช่น ในการเปรียบเทียบผลการเรียนวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนคนหนึ่งว่าเรียนวิชาใดได้ดีกว่ากัน โดยสมมติว่าคะแนนสอบทั้งสองวิชาของนักเรียนในชั้นเรียนนี้มีการแจกแจงปกติ ถ้าพิจารณาจากคะแนนสอบทั้งสองวิชาโดยปรับให้มีคะแนนเต็มเท่ากัน ก็ไม่อาจสรุปได้ว่านักเรียนคนนี้เรียนวิชาใดได้ดีกว่า เนื่องจากคะแนนสอบแต่ละวิชาไม่ได้ขึ้นอยู่กับความรู้ความสามารถในวิชานั้น ๆ ของนักเรียนเพียงอย่างเดียว แต่ยังขึ้นอยู่กับความยากง่ายของข้อสอบหรือวิธีการให้คะแนนของผู้สอนแต่ละวิชา ทำให้ค่าเฉลี่ยหรือส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบทั้งสองวิชาของนักเรียนทั้งหมดในชั้นอาจไม่เท่ากัน ดังนั้น เมื่อต้องการเปรียบเทียบผลการเรียนทั้งสองวิชา จึงจำเป็นต้องแปลงคะแนนสอบทั้งสองวิชาให้เป็นตัวแปรสุ่มปกติมาตรฐาน ซึ่งจะทำให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบทั้งสองวิชาเท่ากัน แล้วจึงเปรียบเทียบจากค่าของตัวแปรสุ่มปกติมาตรฐานของทั้งสองวิชา ดังตัวอย่างต่อไปนี้

ตัวอย่างที่ 17

พิมพ์ชนกสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษซึ่งมีคะแนนเต็ม 100 คะแนนเท่ากัน ได้ 75 และ 72 คะแนน ตามลำดับ ถ้าคะแนนสอบทั้งสองวิชาของนักเรียนห้องนี้มีการแจกแจงปกติ โดยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนห้องนี้เท่ากับ 73 และ 16 คะแนน ตามลำดับ และค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาภาษาอังกฤษของนักเรียนห้องนี้เท่ากับ 70 และ 10 คะแนน ตามลำดับ จงพิจารณาว่าพิมพ์ชนกเรียนวิชาใดได้ดีกว่ากัน

วิธีทำ ให้ตัวแปรสุ่ม X และ Y คือคะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของนักเรียนห้องนี้ ตามลำดับ

จะได้ว่าตัวแปรสุ่ม X และ Y มีการแจกแจงปกติ โดยที่ $\mu_X = 73, \sigma_X = 16$ และ $\mu_Y = 70, \sigma_Y = 10$

ให้ x และ y คือ คะแนนสอบวิชาคณิตศาสตร์และวิชาภาษาอังกฤษของพิมพ์ชนก ตามลำดับ

นั่นคือ $x = 75$ และ $y = 72$

จะได้ว่าค่าของตัวแปรสุ่มปกติมาตรฐานของ x คือ $\dfrac{x-\mu_X}{\sigma_X} = \dfrac{75-73}{16} = 0.125$

และค่าของตัวแปรสุ่มปกติมาตรฐานของ y คือ $\dfrac{y-\mu_Y}{\sigma_Y} = \dfrac{72-70}{10} = 0.2$

เนื่องจาก $\dfrac{x-\mu_X}{\sigma_X} < \dfrac{y-\mu_Y}{\sigma_Y}$

นั่นคือ ค่าของตัวแปรสุ่มปกติมาตรฐานของวิชาคณิตศาสตร์ของพิมพ์ชนกมีค่าน้อยกว่าค่าของตัวแปรสุ่มปกติมาตรฐานของวิชาภาษาอังกฤษของพิมพ์ชนก

ดังนั้น พิมพ์ชนกเรียนวิชาภาษาอังกฤษได้ดีกว่าวิชาคณิตศาสตร์ ■

ตัวอย่างที่ 18

นิติและนิพนธ์เป็นนักเรียนห้องเดียวกันและเข้าสอบวิชาฟิสิกส์ด้วยกัน นิติได้คะแนนสอบ 55 คะแนน ซึ่งปรับเป็นค่าของตัวแปรสุ่มปกติมาตรฐานได้เป็น –0.8 ส่วนนิพนธ์ได้คะแนนสอบ 72 คะแนน ซึ่งปรับเป็นค่าของตัวแปรสุ่มปกติมาตรฐานได้เป็น 1.4 ถ้าคะแนนสอบวิชาฟิสิกส์ของนักเรียนห้องนี้มีการแจกแจงปกติ จงหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาฟิสิกส์ของนักเรียนห้องนี้

วิธีทำ ให้ตัวแปรสุ่ม X คือ คะแนนสอบวิชาฟิสิกส์ของนักเรียนห้องนี้

จะได้ว่าตัวแปรสุ่ม X มีการแจกแจงปกติ โดยมีค่าเฉลี่ย μ และส่วนเบี่ยงเบนมาตรฐาน σ

ให้ x_1 และ x_2 คือ คะแนนสอบวิชาฟิสิกส์ของนิติและนิพนธ์ ตามลำดับ

นั่นคือ $x_1 = 55$ และ $x_2 = 72$

จะได้ว่าค่าของตัวแปรสุ่มปกติมาตรฐานของ x_1 คือ $\frac{55-\mu}{\sigma} = -0.8$

นั่นคือ $\mu = 55 + 0.8\sigma$

และค่าของตัวแปรสุ่มปกติมาตรฐานของ x_2 คือ $\frac{72-\mu}{\sigma} = 1.4$

นั่นคือ $\mu = 72 - 1.4\sigma$

จะได้ $55 + 0.8\sigma = 72 - 1.4\sigma$

ดังนั้น $\sigma \approx 7.73$ และ $\mu \approx 61.18$

สรุปได้ว่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาฟิสิกส์ของนักเรียนห้องนี้มีค่าประมาณ 61.18 และ 7.73 คะแนน ตามลำดับ ■

จากตัวอย่างข้างต้นจะสังเกตได้ว่า ค่าของตัวแปรสุ่มปกติมาตรฐานอาจเป็นได้ทั้งจำนวนจริงบวกและจำนวนจริงลบ ถ้าค่าของตัวแปรสุ่มปกติมาตรฐานเป็นจำนวนจริงบวก แสดงว่าข้อมูลนั้นมีค่ามากกว่าค่าเฉลี่ย แต่ถ้าค่าของตัวแปรสุ่มปกติมาตรฐานเป็นจำนวนจริงลบ แสดงว่าข้อมูลนั้นมีค่าน้อยกว่าค่าเฉลี่ย นอกจากนี้ ค่าของตัวแปรสุ่มปกติมาตรฐานจะขึ้นอยู่กับส่วนเบี่ยงเบนมาตรฐานด้วย กล่าวคือถ้าส่วนเบี่ยงเบนมาตรฐานมีค่าน้อย จะทำให้ค่าของตัวแปรสุ่มปกติมาตรฐานมีค่ามากกว่ากรณีที่ส่วนเบี่ยงเบนมาตรฐานมีค่ามาก

แบบฝึกหัด 4.3

1. ให้ Z เป็นตัวแปรสุ่มปกติมาตรฐาน จงหา
   1) $P(Z \le -1.34)$
   2) $P(Z > 2.18)$
   3) $P(-2.45 \le Z \le 1.68)$
   4) $P(0.91 \le Z \le 2.26)$

2. ข้อมูลชุดหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 และ 2 ตามลำดับ จงหาค่าของตัวแปรสุ่มปกติมาตรฐานของข้อมูลต่อไปนี้
   1) 11
   2) 12
   3) 14.2

3. กำหนดให้ $X \sim N(20, 100)$ จงหา
   1) $P(X < 48.9)$
   2) $P(X > 12.9)$
   3) $P(18.5 \le X \le 37.4)$

4. ข้อมูลชุดหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 400 และ 100 ตามลำดับ จงหาว่ามีข้อมูลอยู่กี่เปอร์เซ็นต์ที่มีค่า
   1) มากกว่า 538
   2) มากกว่า 179
   3) น้อยกว่า 356
   4) น้อยกว่า 621

5. ระยะเวลาในการประกอบชิ้นส่วนอิเล็กทรอนิกส์ชนิดหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 12 และ 1.5 นาที ตามลำดับ ถ้าสุ่มชิ้นส่วนอิเล็กทรอนิกส์มา 1 ชิ้น จงหาความน่าจะเป็นที่ชิ้นส่วนอิเล็กทรอนิกส์นี้จะใช้เวลาในการประกอบ
   1) น้อยกว่า 11 นาที
   2) ระหว่าง 10 และ 13 นาที
   3) มากกว่า 14 นาที

6. คะแนนสอบวัดความรู้ความสามารถวิชาคณิตศาสตร์ของนักเรียนระดับมัธยมศึกษาตอนปลายที่จัดโดยสถาบันแห่งหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 505 และ 111 คะแนน ตามลำดับ ถ้าสุ่มนักเรียน 1 คน ที่เข้าร่วมการสอบนี้ จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้คะแนนสอบ
   1) ระหว่าง 400 และ 600 คะแนน
   2) มากกว่า 700 คะแนน
   3) น้อยกว่า 450 คะแนน

7. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 72 และ 12 คะแนน ตามลำดับ ถ้ากำหนดให้ $P(Z < -0.6745) = 0.25$ และ $P(Z < 1.2816) = 0.90$ จงหาเปอร์เซ็นไทล์ที่ 25 และเปอร์เซ็นไทล์ที่ 90

8. ในการผลิตแผ่นพลาสติกของบริษัทแห่งหนึ่ง พบว่า ความหนาของแผ่นพลาสติกมีการแจกแจงปกติ โดยมีความหนาเฉลี่ย 0.0625 เซนติเมตร และความแปรปรวน 0.00000625 เซนติเมตร^2 ถ้าสุ่มแผ่นพลาสติกมา 1 แผ่น จงหาความน่าจะเป็นที่แผ่นพลาสติกที่สุ่มได้มีความหนามากกว่า 0.0595 เซนติเมตร แต่ไม่เกิน 0.0659 เซนติเมตร

9. คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 22 และ 4 คะแนน ตามลำดับ
   1) จงหาค่าของตัวแปรสุ่มปกติมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนที่ได้ 30 คะแนน
   2) ถ้าสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 มา 1 คน จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้คะแนนสอบระหว่าง 15 และ 32 คะแนน
   3) ถ้าสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 มา 1 คน จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้คะแนนสอบมากกว่า 34 คะแนน
   4) ถ้าสุ่มนักเรียนชั้นมัธยมศึกษาปีที่ 6 มา 1 คน จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้คะแนนสอบน้อยกว่า 25 คะแนน

10. น้ำหนักสุทธิของกระป๋องบรรจุถั่วที่ผลิตในโรงงานแห่งหนึ่งมีการแจกแจงปกติ โดยมีน้ำหนักสุทธิเฉลี่ย 12 กรัม ถ้าสุ่มกระป๋องบรรจุถั่วมา 1 ใบ แล้วความน่าจะเป็นที่กระป๋องใบนี้จะมีน้ำหนักสุทธิน้อยกว่า 11.8 กรัม คือ 0.1151 จงหาความแปรปรวนของน้ำหนักสุทธิของกระป๋องบรรจุถั่วที่ผลิตโดยโรงงานแห่งนี้

11. ร้านอาหารแห่งหนึ่งนอกจากขายอาหารแล้วยังมีผลไม้ขายบริเวณหน้าร้านด้วย โดยยอดขายอาหารรายวันมีการแจกแจงปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 8,400 และ 360 บาท ตามลำดับ และยอดขายผลไม้รายวันมีการแจกแจงปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5,200 และ 240 บาท ตามลำดับ ถ้าวันนี้มียอดขายอาหาร 9,500 บาท และยอดขายผลไม้ 6,000 บาท จงพิจารณาว่าวันนี้ร้านอาหารแห่งนี้ขายอาหารหรือผลไม้ได้ดีกว่ากัน

12. ภัครพรรณต้องการเรียนต่อคณะอักษรศาสตร์ ณ มหาวิทยาลัยแห่งหนึ่ง จึงได้สมัครสอบวัดระดับความถนัดทางภาษาจีนและภาษาเกาหลี และได้คะแนนสอบดังนี้

ภาษา	คะแนนของภัครพรรณ	คะแนนเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
จีน	75	55	13
เกาหลี	68	50	10

ถ้าคะแนนสอบวัดระดับความถนัดทางภาษาจีนและภาษาเกาหลีมีการแจกแจงปกติ แล้วภัครพรรณมีโอกาสสอบติดคณะอักษรศาสตร์สาขาวิชาภาษาจีนหรือสาขาวิชาภาษาเกาหลีมากกว่ากัน เพราะเหตุใด

13. เกณฑ์ในการแบ่งระดับคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของโรงเรียนแห่งหนึ่ง แสดงได้ดังนี้

เกรด	คะแนน
4	ตั้งแต่ $\mu+1.5\sigma$ ขึ้นไป
3	$[\mu+0.5\sigma, \mu+1.5\sigma)$
2	$[\mu-0.5\sigma, \mu+0.5\sigma)$
1	$[\mu-1.5\sigma, \mu-0.5\sigma)$
0	น้อยกว่า $\mu-1.5\sigma$

โดย μ และ σ แทนค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ตามลำดับ สมมติว่าคะแนนสอบมีการแจกแจงปกติ ถ้าสุ่มนักเรียนมา 1 คน จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้เกรดในแต่ละระดับ

14. อักษรศิลป์และสรวิทย์เป็นนักเรียนห้องเดียวกัน อักษรศิลป์สูง 152 เซนติเมตร ซึ่งปรับเป็นค่าของตัวแปรสุ่มปกติมาตรฐานได้เป็น -1.2 ส่วนสรวิทย์สูง 170 เซนติเมตร ซึ่งปรับเป็นค่าของตัวแปรสุ่มปกติมาตรฐานได้เป็น 0.8 ถ้าส่วนสูงของนักเรียนห้องนี้มีการแจกแจงปกติ
   1) จงหาค่าเฉลี่ยและความแปรปรวนของความสูงของนักเรียนห้องนี้
   2) ถ้าวิภารัตน์สูง 160 เซนติเมตร วิภารัตน์จะสูงกว่าค่าเฉลี่ยของความสูงของนักเรียนห้องนี้หรือไม่
   3) ถ้าเกริกสูง 180 เซนติเมตร จงหาความน่าจะเป็นที่นักเรียนคนหนึ่งที่สุ่มมาจากห้องนี้จะสูงกว่าเกริก

กิจกรรม : นอนพอไหม

นักวิจัยได้สำรวจจำนวนชั่วโมงการนอนในคืนหนึ่งของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในอำเภอเมืองชลบุรี จังหวัดชลบุรี โดยเลือกตัวอย่างนักเรียนจากโรงเรียนต่าง ๆ ในอำเภอจำนวน 132 คน ได้ผลสำรวจดังนี้

6.0	7.0	5.0	7.5	8.5	9.0	7.0	10.5	4.0	11.5	8.0	7.5
8.5	4.0	8.0	3.5	9.0	5.0	6.0	5.5	10.5	9.5	12.0	5.5
5.5	11.0	6.5	9.0	7.0	9.0	4.0	7.5	8.5	4.5	7.0	10.5
9.5	4.5	8.5	7.5	5.5	9.0	10.0	3.0	8.0	6.5	10.0	6.5
10.0	4.5	10.5	6.5	5.0	4.0	7.5	9.5	6.5	7.5	5.0	8.0
6.5	10.0	7.0	7.5	6.0	8.5	8.0	6.0	10.5	6.5	9.5	8.5
8.5	9.5	6.5	8.0	7.0	3.5	7.5	4.0	5.5	9.0	3.5	11.0
7.0	5.0	3.0	6.5	4.5	10.0	8.0	6.0	7.0	8.0	7.5	6.0
7.5	6.0	9.0	8.0	7.5	5.5	4.5	6.0	8.5	5.0	7.0	9.5
7.0	7.5	5.5	8.5	7.0	6.5	8.0	7.0	5.5	9.0	6.0	11.5
6.0	8.0	5.0	7.5	6.5	8.5	9.5	5.0	10.0	11.0	4.5	7.0

นักเรียนสามารถคัดลอกชุดข้อมูลนี้ได้ที่ ipst.me/10681

สมมติว่าตัวอย่างที่เลือกมานี้เป็นตัวแทนที่ดีของประชากร และจำนวนชั่วโมงการนอนของนักเรียนมีการแจกแจงใกล้เคียงกับการแจกแจงปกติ

ให้นักเรียนตอบคำถามต่อไปนี้

1. ระบุประชากรและตัวอย่างในการสำรวจครั้งนี้
2. กำหนดตัวแปรสุ่มที่ใช้ในการสำรวจครั้งนี้และระบุชนิดของตัวแปรสุ่ม พร้อมทั้งให้เหตุผลประกอบ
3. หาค่าเฉลี่ยเลขคณิตและส่วนเบี่ยงเบนมาตรฐานของข้อมูลชุดนี้ พร้อมทั้งอธิบายความหมาย
4. เนื่องจากนักวิจัยผู้นี้ได้เลือกตัวอย่างที่เป็นตัวแทนที่ดีของประชากร ดังนั้นจะกำหนดให้ค่าเฉลี่ยเลขคณิตของตัวอย่างเป็นค่าเฉลี่ยของประชากร (μ) และกำหนดให้ส่วนเบี่ยงเบนมาตรฐานของตัวอย่างเป็นส่วนเบี่ยงเบนมาตรฐานของประชากร (σ) และเนื่องจากจำนวนชั่วโมงการนอนของนักเรียนมีการแจกแจงใกล้เคียงกับการแจกแจงปกติ จงใช้โปรแกรม GeoGebra สร้างเส้นโค้งปกติแสดงจำนวนชั่วโมงการนอนของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในอำเภอเมืองชลบุรี จังหวัดชลบุรี
5. National Sleep Foundation ของสหรัฐอเมริกาได้แนะนำระยะเวลาในการนอนที่เหมาะสมของวัยรุ่นอายุ 14 – 17 ปี คือ 8 – 10 ชั่วโมงต่อคืน จงใช้เส้นโค้งปกติที่ได้จากข้อ 4 หาร้อยละของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในอำเภอเมืองชลบุรี จังหวัดชลบุรี ที่มีจำนวนชั่วโมงการนอนไม่สอดคล้องกับคำแนะนำของ National Sleep Foundation

แบบฝึกหัดท้ายบท

1. จงพิจารณาว่าตัวแปรสุ่มต่อไปนี้เป็นตัวแปรสุ่มไม่ต่อเนื่องหรือตัวแปรสุ่มต่อเนื่อง
   1) ตัวแปรสุ่ม X_1 คือ จำนวนต้นไม้ที่ชาวสวนคนหนึ่งปลูกได้ในเวลา 1 เดือน
   2) ตัวแปรสุ่ม X_2 คือ ความสูงของผู้ป่วยในโรงพยาบาลแห่งหนึ่ง
   3) ตัวแปรสุ่ม X_3 คือ จำนวนนักเรียนที่ชื่นชอบภาพยนตร์สยองขวัญ จากการสอบถามนักเรียนชั้นมัธยมศึกษาตอนปลายจำนวน 100 คน
   4) ตัวแปรสุ่ม X_4 คือ น้ำหนักของเงาะกระป๋องที่ผลิตในโรงงานแห่งหนึ่ง
   5) ตัวแปรสุ่ม X_5 คือ เวลาที่ใช้ในการเดินทางจากจังหวัดเชียงใหม่ถึงจังหวัดอุดรธานีโดยเครื่องบิน
   6) ตัวแปรสุ่ม X_6 คือ ปริมาณการใช้น้ำประปาของโรงเรียนแห่งหนึ่งใน 1 วัน
   7) ตัวแปรสุ่ม X_7 คือ ผลคูณของแต้มบนหน้าลูกเต๋าจากการทอดลูกเต๋า 2 ลูก พร้อมกัน 1 ครั้ง

2. จำนวนวันที่นักเรียนชั้นมัธยมศึกษาปีที่ 6 ห้องหนึ่งออกกำลังกายใน 1 สัปดาห์ แสดงด้วยตารางความถี่ได้ดังนี้

จำนวนวันที่นักเรียนออกกำลังกายใน 1 สัปดาห์	จำนวนนักเรียน (คน)
0	1
1	7
2	10
3	12
4	6
5	3
6	4
7	1

ถ้าสุ่มนักเรียน 1 คน จากห้องนี้ และให้ตัวแปรสุ่ม X คือจำนวนวันที่นักเรียนที่สุ่มได้ออกกำลังกายใน 1 สัปดาห์

1) จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตาราง
2) จงหาค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

3. ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่ลูกเต๋าขึ้นแต้ม 3 หรือ 4 จากการทอดลูกเต๋าที่เที่ยงตรง 1 ลูก 2 ครั้ง

1) จงเขียนแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X ในรูปตาราง
2) จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

4. สลากกินแบ่งรัฐบาล 1 ฉบับ ราคา 80 บาท แต่ละฉบับประกอบด้วยเลขโดด 6 ตัว โดยมีเงื่อนไขการตรวจผลรางวัลดังนี้

รางวัล	จำนวน	มูลค่า (บาท)
รางวัลที่ หนึ่ง	1 รางวัล	6,000,000
รางวัลที่ สอง	5 รางวัล	200,000
รางวัลที่ สาม	10 รางวัล	80,000
รางวัลที่ สี่	50 รางวัล	40,000
รางวัลที่ ห้า	100 รางวัล	20,000
รางวัลข้างเคียงรางวัลที่หนึ่ง	2 รางวัล	100,000
รางวัลเลขหน้า 3 ตัว เสี่ยง 2 ครั้ง	2,000 รางวัล	4,000
รางวัลเลขท้าย 3 ตัว เสี่ยง 2 ครั้ง	2,000 รางวัล	4,000
รางวัลเลขท้าย 2 ตัว เสี่ยง 1 ครั้ง	10,000 รางวัล	2,000

ถ้าซื้อสลากกินแบ่งรัฐบาล 1 ฉบับ จะมีโอกาสได้กำไรหรือขาดทุนมากกว่ากัน โดยสมมติว่าจะสามารถรับรางวัลได้เพียงรางวัลเดียว

5. กล่องใบหนึ่งบรรจุสลาก 10 ใบ โดยมีหมายเลข 1, 2, 3, … , 10 กำกับไว้ ถ้านักเรียนสุ่มหยิบสลากจากกล่องใบนี้ 1 ใบ และให้ตัวแปรสุ่ม X คือหมายเลขบนสลากที่สุ่มได้
   1) จงพิจารณาว่าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงเอกรูปไม่ต่อเนื่องหรือไม่
   2) จงหาค่าคาดหมายและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

6. ให้ตัวแปรสุ่ม Y มีค่าที่เป็นไปได้เป็นจำนวนนับที่เรียงติดกันทั้งหมด 7 ค่า โดยมีมัธยฐานเป็น 10 ถ้าการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม Y เป็นการแจกแจงเอกรูปไม่ต่อเนื่อง จงหาค่าคาดหมาย ความแปรปรวน และส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม Y

7. กำหนดให้ $X \sim B(8, 0.7)$ และค่าที่เป็นไปได้ของตัวแปรสุ่ม X คือ 0, 1, 2, 3, 4, 5, 6, 7, 8 จงหา
   1) $P(X = 4)$
   2) $P(X \geq 6)$
   3) $P(X > 5)$
   4) $P(3 \leq X \leq 7)$

8. สามีภรรยาคู่หนึ่งต้องการมีลูก 4 คน ถ้าสมมติว่าโอกาสที่ลูกแต่ละคนจะเป็นชายหรือหญิงเท่ากัน และให้ตัวแปรสุ่ม X คือจำนวนลูกชายของสามีภรรยาคู่นี้ จงหา
   1) ค่าคาดหมายของตัวแปรสุ่ม X พร้อมทั้งอธิบายความหมาย
   2) ความแปรปรวนและส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม X

9. นักฟุตบอลคนหนึ่งมีโอกาสยิงเข้าประตู 70% ในการยิงประตูแต่ละครั้ง ถ้านักฟุตบอลคนนี้ยิงประตู 5 ครั้ง จงหาความน่าจะเป็นที่
   1) เขาจะยิงได้ 3 ประตู
   2) เขาจะยิงได้อย่างน้อย 2 ประตู
   3) เขาจะยิงไม่เข้าประตูเลย

10. จากข้อมูลในอดีตของชุมชนแห่งหนึ่ง พบว่า ความน่าจะเป็นที่แต่ละคนในชุมชนจะเป็นโรคเบาหวานเท่ากับ 0.1 ถ้าสุ่มคนในชุมชนแห่งนี้มา 25 คน
    1) ความน่าจะเป็นที่จะมีอย่างน้อย 2 คน ในกลุ่มคนที่สุ่มมาเป็นโรคเบาหวานเป็นเท่าใด
    2) คาดว่าจะมีกี่คนในกลุ่มคนที่สุ่มมาเป็นโรคเบาหวาน

11. ลักษณะหมู่เลือดระบบ ABO เป็นลักษณะทางพันธุกรรมที่มีแอลลีลที่เกี่ยวข้อง 3 รูปแบบ คือ I^A, I^B และ i ในกรณีที่พ่อมีเลือดหมู่ A $\left(I^A i\right)$ และแม่มีเลือดหมู่ B $\left(I^B i\right)$ โอกาสที่จะได้ลูกซึ่งมีเลือดหมู่ต่าง ๆ มีดังนี้

    โอกาสที่ลูกจะมีเลือดหมู่ A $\left(I^A i\right)$ เป็น $\frac{1}{4}$

    โอกาสที่ลูกจะมีเลือดหมู่ B $\left(I^B i\right)$ เป็น $\frac{1}{4}$

    โอกาสที่ลูกจะมีเลือดหมู่ AB $\left(I^A I^B\right)$ เป็น $\frac{1}{4}$

    โอกาสที่ลูกจะมีเลือดหมู่ O (ii) เป็น $\frac{1}{4}$

    ถ้าสามีภรรยาคู่หนึ่งต้องการมีลูก 4 คน โดยสามีมีเลือดหมู่ A $\left(I^A i\right)$ และภรรยามีเลือดหมู่ B $\left(I^B i\right)$ จงหา
    1) ความน่าจะเป็นที่จะมีลูกมากกว่า 2 คน มีเลือดหมู่ A $\left(I^A i\right)$
    2) ความน่าจะเป็นที่ลูกทั้งสี่คนไม่มีเลือดหมู่ O (ii)

12. จากข้อมูลเกี่ยวกับคุณภาพของผลิตภัณฑ์ซึ่งเก็บรวบรวมจากโรงงานแห่งหนึ่งทำให้ทราบว่า ความน่าจะเป็นที่สินค้าแต่ละกล่องที่ผลิตจากโรงงานแห่งหนึ่งจะมีผลิตภัณฑ์ที่ฝาเกลียวเปิดยากเท่ากับ 0.04 และในกระบวนการตรวจสอบคุณภาพผลิตภัณฑ์ของโรงงาน มีหลักการคือพนักงานจะสุ่มผลิตภัณฑ์จำนวน 5 ชิ้น จากแต่ละกล่องเพื่อตรวจสอบคุณภาพ
    1) จงหาความน่าจะเป็นที่ผลิตภัณฑ์ที่สุ่มมาจากแต่ละกล่องเป็นผลิตภัณฑ์ที่ฝาเกลียวเปิดยากทั้ง 5 ชิ้น
    2) ถ้าตรวจพบผลิตภัณฑ์ที่ฝาเกลียวเปิดยากไม่เกิน 2 ชิ้น จากผลิตภัณฑ์ที่สุ่มมา ผลิตภัณฑ์กล่องนั้นจะผ่านการตรวจสอบคุณภาพ จงหาความน่าจะเป็นที่ผลิตภัณฑ์แต่ละกล่องที่ส่งมาตรวจสอบจะผ่านการตรวจสอบคุณภาพ (ตอบเป็นทศนิยม 4 ตำแหน่ง)

3) ถ้าฝ่ายผลิตของโรงงานส่งผลิตภัณฑ์มาให้พนักงานตรวจสอบคุณภาพจำนวน 10,000 กล่อง จะมีผลิตภัณฑ์ที่ไม่ผ่านการตรวจสอบคุณภาพประมาณกี่กล่อง

13. ความน่าจะเป็นที่เครื่องขยายเสียงแต่ละเครื่องที่ผลิตจากบริษัทแห่งหนึ่งจะมีระดับเสียงมากกว่า 90 เดซิเบล เท่ากับ 0.05 ถ้าสุ่มเครื่องขยายเสียงจากบริษัทแห่งนี้เพื่อวัดระดับเสียงจำนวน 12 เครื่อง จงหาความน่าจะเป็นที่มีเครื่องขยายเสียงอย่างมาก 2 เครื่อง ที่มีระดับเสียงมากกว่า 90 เดซิเบล

14. เกมวงล้อเสี่ยงโชคมีกติกาการเล่นคือ ผู้เล่นจะต้องหมุนวงล้อรูปวงกลมที่แบ่งเป็น 6 ช่องเท่า ๆ กัน โดยแต่ละช่องระบุจำนวนเงินรางวัลแตกต่างกันคือ 50, 100, 200, 300, 400 และ 500 บาท ถ้าลูกศรชี้ที่ช่องใด ผู้เล่นจะได้เงินรางวัลตามที่ระบุในช่องนั้น ให้ตัวแปรสุ่ม X คือเงินรางวัลที่ได้จากการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง สมมติในการหมุนวงล้อแต่ละครั้งโอกาสที่ลูกศรจะชี้ที่ช่องใดช่องหนึ่งเท่ากัน

1) จงพิจารณาว่าตัวแปรสุ่ม X มีการแจกแจงแบบใด พร้อมทั้งให้เหตุผลประกอบ
2) โดยเฉลี่ยแล้วในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะได้เงินรางวัลกี่บาท และส่วนเบี่ยงเบนมาตรฐานของเงินรางวัลที่ได้จากการเล่มเกมวงล้อเสี่ยงโชคแต่ละครั้งเป็นเท่าใด
3) ถ้าในการเล่นเกมวงล้อเสี่ยงโชคแต่ละครั้ง ผู้เล่นจะต้องจ่ายเงินซื้อตั๋วราคา 250 บาท นักเรียนจะตัดสินใจเล่นเกมนี้หรือไม่ เพราะเหตุใด

15. ให้ตัวแปรสุ่ม X คือจำนวนครั้งที่ลูกเต๋าขึ้นแต้มมากกว่า 4 จากการทอดลูกเต๋าที่เที่ยงตรง 1 ลูก 10 ครั้ง

1) จงพิจารณาว่าตัวแปรสุ่ม X มีการแจกแจงแบบใด พร้อมทั้งให้เหตุผลประกอบ
2) จงหาความน่าจะเป็นที่ลูกเต๋าขึ้นแต้มมากกว่า 4 เป็นจำนวน 4 ถึง 6 ครั้ง
3) จงหาค่าคาดหมายและความแปรปรวนของตัวแปรสุ่ม X

16 ข้อสอบปรนัยชุดหนึ่งมี 10 ข้อ ข้อละ 1 คะแนน โดยแต่ละข้อมี 6 ตัวเลือก และมีคำตอบที่ถูกต้องเพียง 1 คำตอบเท่านั้น ถ้านักเรียนคนหนึ่งทำข้อสอบชุดนี้ด้วยการเดาทุกข้อ และให้ตัวแปรสุ่ม X คือคะแนนที่นักเรียนคนนี้ได้รับจากการทำข้อสอบชุดนี้

1) จงพิจารณาว่าตัวแปรสุ่ม X มีการแจกแจงแบบใด พร้อมทั้งให้เหตุผลประกอบ
2) จงหาความน่าจะเป็นที่นักเรียนคนนี้จะตอบผิดทุกข้อ
3) ถ้าเกณฑ์ในการสอบผ่านคือต้องได้ตั้งแต่ 5 คะแนนขึ้นไป จงหาความน่าจะเป็นที่นักเรียนคนนี้จะได้คะแนนผ่านเกณฑ์พอดีและความน่าจะเป็นที่นักเรียนคนนี้จะสอบผ่าน (ตอบเป็นทศนิยม 3 ตำแหน่ง)
4) โดยเฉลี่ยแล้วนักเรียนคนนี้จะทำข้อสอบชุดนี้ได้กี่คะแนน และความแปรปรวนของคะแนนที่นักเรียนคนนี้ได้รับจากการทำข้อสอบชุดนี้เป็นเท่าใด

17 กำหนดให้ $X \sim N(100, 64)$ จงหา

1) $P(X < 92)$ 2) $P(94 < X < 106)$ 3) $P(102 < X < 110)$

18 กำหนดให้ $X \sim N(25, 36)$ จงหา $P(20 < X < 35)$

19 กำหนดให้ $X \sim N(120, 225)$ จงหา $P(X \geq 130)$ และ $P(93 \leq X \leq 140)$

20 น้ำหนักสุทธิของกาแฟผงสำเร็จรูปบรรจุขวดที่ผลิตในโรงงานแห่งหนึ่งมีการแจกแจงปกติ โดยมีน้ำหนักสุทธิเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 115.5 และ 0.3 กรัม ตามลำดับ ถ้าสุ่มกาแฟผงสำเร็จรูปมา 1 ขวด จงหาความน่าจะเป็นที่กาแฟผงสำเร็จรูปบรรจุขวดที่สุ่มได้จะมีน้ำหนักสุทธิ

1) ระหว่าง 115 และ 115.5 กรัม
2) ระหว่าง 114.9 และ 115.5 กรัม
3) ระหว่าง 115.2 และ 115.9 กรัม
4) ระหว่าง 114.7 และ 115 กรัม
5) มากกว่า 115.5 กรัม
6) น้อยกว่า 115 กรัม

21 สำนักวิจัยแห่งหนึ่งรายงานว่าค่าใช้จ่ายในการใช้โทรศัพท์มือถือรายเดือนของคนในจังหวัดหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 520 และ 270 บาท ตามลำดับ ถ้าสุ่มคนจากจังหวัดนี้มา 1 คน จงหาความน่าจะเป็นที่ผู้ที่สุ่มได้จะมีค่าใช้จ่ายในการใช้โทรศัพท์มือถือ

1) น้อยกว่า 520 บาทต่อเดือน
2) มากกว่า 700 บาทต่อเดือน
3) ระหว่าง 500 และ 800 บาทต่อเดือน

22 ในการสอบครั้งหนึ่งมีนักเรียนเข้าสอบ 1,000 คน โดยคะแนนสอบมีการแจกแจงปกติที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 200 และ 30 คะแนน ตามลำดับ จงหาว่ามีนักเรียนประมาณกี่คนที่ได้คะแนนสอบ

1) ระหว่าง 170 และ 230 คะแนน
2) มากกว่า 260 คะแนน

23 คะแนนสอบวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในภาคเรียนหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 60 และ 5 คะแนน ตามลำดับ ถ้ากำหนดให้ $P(Z<1.645)=0.95$ และ $P(Z<-1.282)=0.1$ จงหา

1) คะแนนต่ำสุดของกลุ่มนักเรียนที่ได้คะแนนสูงสุด ซึ่งมีจำนวนประมาณ 5% ของนักเรียนทั้งหมด
2) คะแนนสูงสุดของกลุ่มนักเรียนที่ได้คะแนนต่ำสุด ซึ่งมีจำนวนประมาณ 10% ของนักเรียนทั้งหมด

24 คะแนนสอบวัดผลวิชาคณิตศาสตร์ระดับประเทศที่จัดโดยสถาบันแห่งหนึ่งมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 25 และ 5 คะแนน ตามลำดับ ถ้าสุ่มผู้เข้าสอบมา 1 คน และกำหนดให้ $P(Z<2.054)=0.98$ จงหา

1) ความน่าจะเป็นที่ผู้เข้าสอบที่สุ่มได้จะได้คะแนนสอบมากกว่า 30 คะแนน
2) คะแนนต่ำสุดของกลุ่มผู้เข้าสอบที่ได้คะแนนสูงสุด ซึ่งมีจำนวนประมาณ 2% ของผู้เข้าสอบทั้งหมด

25 จากการบันทึกข้อมูลน้ำหนักรถรวมน้ำหนักบรรทุกของรถบรรทุกที่วิ่งระหว่างกรุงเทพมหานครและนครราชสีมาในช่วงระยะเวลาหนึ่ง พบว่า ข้อมูลดังกล่าวมีการแจกแจงปกติ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 15 และ 4 ตัน ตามลำดับ ถ้าสุ่มรถบรรทุกที่วิ่งเส้นทางนี้มา 1 คัน จงหาความน่าจะเป็นที่รถบรรทุกที่สุ่มได้จะมีน้ำหนักรถรวมน้ำหนักบรรทุก

1) มากกว่า 12 ตัน
2) น้อยกว่า 20 ตัน
3) ระหว่าง 10 และ 18 ตัน

เสริมสมอง : พิกัดน้ำหนักรถบรรทุกตามที่กฎหมายกำหนด

กฎหมายกำหนดให้น้ำหนักรถรวมน้ำหนักบรรทุกสำหรับรถบรรทุกแต่ละประเภทต้องไม่เกินน้ำหนักรวมสูงสุดดังต่อไปนี้

รถบรรทุก 2 เพลา 4 ล้อ น้ำหนักรวมสูงสุด 9.5 ตัน
รถบรรทุก 2 เพลา 6 ล้อ น้ำหนักรวมสูงสุด 15 ตัน
รถบรรทุก 3 เพลา 10 ล้อ น้ำหนักรวมสูงสุด 25 ตัน

หากฝ่าฝืนจะมีความผิดตามพระราชบัญญัติทางหลวง พ.ศ. 2535 มาตรา 61 ต้องระวางโทษจำคุกไม่เกิน 6 เดือน หรือปรับไม่เกิน 10,000 บาท หรือทั้งจำทั้งปรับ

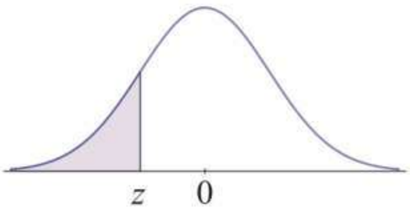
26 กำหนดให้ $X \sim N(\mu, \sigma^2)$ จงหา a, b, c และ d จากข้อมูลที่กำหนดให้ต่อไปนี้

μ	σ	x	$P(X \le x)$
3	1	2	d
10	2	c	0.1841
a	3	6	0.0985
10	b	12	0.6026

27 เก่งและกล้าเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 โรงเรียนเดียวกันแต่อยู่คนละห้อง โดยเก่งอยู่ห้อง 1 และกล้าอยู่ห้อง 2 ซึ่งห้องเรียนแต่ละห้องประกอบด้วยนักเรียนที่มีความสามารถแตกต่างกัน ถ้าเก่งและกล้าเรียนวิชาคณิตศาสตร์ที่สอนโดยครูคนละคนกันและใช้ข้อสอบแตกต่างกัน โดยในการสอบกลางภาค เก่งและกล้าได้ 80 และ 90 คะแนน ตามลำดับ สมมติว่าคะแนนสอบกลางภาควิชาคณิตศาสตร์ของนักเรียนทั้งสองห้องมีการแจกแจงปกติ โดยคะแนนสอบของนักเรียนห้อง 1 มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 65 และ 5 คะแนน ตามลำดับ และคะแนนสอบของนักเรียนห้อง 2 มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 70 และ 10 คะแนน ตามลำดับ

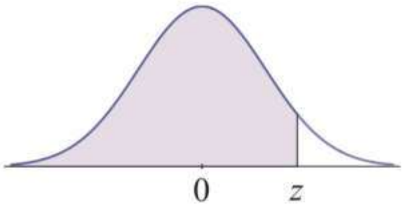
1) มีนักเรียนห้อง 1 กี่เปอร์เซ็นต์ที่ได้คะแนนสอบระหว่าง 55 และ 75 คะแนน
2) มีนักเรียนห้อง 1 กี่เปอร์เซ็นต์ที่ได้คะแนนไม่เกินคะแนนของเก่ง
3) มีนักเรียนห้อง 2 กี่เปอร์เซ็นต์ที่ได้คะแนนมากกว่ากล้า
4) สามารถสรุปได้หรือไม่ว่านักเรียนห้อง 1 มีโอกาสสอบได้คะแนนระหว่าง 55 และ 75 คะแนน มากกว่านักเรียนห้อง 2 มีโอกาสสอบได้คะแนนระหว่าง 55 และ 75 คะแนน
5) จงหาค่าของตัวแปรสุ่มปกติมาตรฐานของคะแนนสอบของเก่งและกล้า และจากค่าของตัวแปรสุ่มปกติมาตรฐาน สามารถบอกได้หรือไม่ว่าใครทำคะแนนสอบได้ดีกว่ากัน

ตารางที่ 1 แสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

ตารางที่ 1 แสดงพื้นที่ใต้เส้นโค้งปกติมาตรฐาน (ต่อ)



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

บรรณานุกรม

เมดไทย (Medthai). (2561, 27 มกราคม). *การตรวจน้ำตาลในเลือด / ตรวจเบาหวาน (Fasting Blood Sugar : FBS / FPG)*. สืบค้นเมื่อ 1 สิงหาคม 2562, จาก https://medthai.com/การตรวจน้ำตาลในเลือด

มูลนิธิโลกสีเขียว. *ศัพท์สิ่งแวดล้อม: อังกฤษ*. สืบค้นเมื่อ 17 มิถุนายน 2562, จาก https://greenworld.or.th/vocab_category_en/g-h-i

มูลนิธิหัวใจแห่งประเทศไทย ในพระบรมราชูปถัมภ์. (2560, 25 ตุลาคม). *คอเลสเตอรอล นางเอก vs นางร้าย*. สืบค้นเมื่อ 20 มีนาคม 2562, จาก http://www.thaiheartfound.org/category/details/food/245

ศุภทัต ชุมนุมวัฒน์. (2559, 23 กันยายน). *ไขมันในเลือดสูงกับโรคหลอดเลือด*. สืบค้นเมื่อ 20 มีนาคม 2562, จาก https://www.pharmacy.mahidol.ac.th/th/knowledge/article/349/ไขมันในเลือดสูง

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2559). *หนังสือเรียนรายวิชาพื้นฐานคณิตศาสตร์ เล่ม 3 ชั้นมัธยมศึกษาปีที่ 4–6 กลุ่มสาระการเรียนรู้คณิตศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551* (พิมพ์ครั้งที่ 10). กรุงเทพฯ: โรงพิมพ์ สกสค. ลาดพร้าว.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2561). *หนังสือเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ เล่ม 5 ชั้นมัธยมศึกษาปีที่ 4–6 กลุ่มสาระการเรียนรู้คณิตศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551* (พิมพ์ครั้งที่ 11). กรุงเทพฯ: โรงพิมพ์ สกสค. ลาดพร้าว.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2561). *หนังสือเรียนรายวิชาพื้นฐานคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เล่ม 2 ตามมาตรฐานการเรียนรู้และตัวชี้วัด กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพฯ: โรงพิมพ์ สกสค. ลาดพร้าว.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2561). *หนังสือเรียนรายวิชาพื้นฐานคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เล่ม 2 ตามมาตรฐานการเรียนรู้และตัวชี้วัด กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2562). *หนังสือเรียนรายวิชาเพิ่มเติมคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 5 เล่ม 2 ตามผลการเรียนรู้ กลุ่มสาระการเรียนรู้คณิตศาสตร์ (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.* กรุงเทพฯ: ศูนย์หนังสือแห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. *พจนานุกรมศัพท์วิทยาศาสตร์ คณิตศาสตร์ และเทคโนโลยี: พาณิชย์อิเล็กทรอนิกส์.* สืบค้นเมื่อ 24 มกราคม 2562, จาก https://escivocab.ipst.ac.th/vocab/พาณิชย์อิเล็กทรอนิกส์

สำนักงานควบคุมน้ำหนักยานพาหนะ กรมทางหลวง. *พิกัดน้ำหนักรถบรรทุกตามที่กฎหมายกำหนด.* สืบค้นเมื่อ 19 สิงหาคม 2562, จาก http://www.highwayweigh.go.th/images/pdf/bosoure/bosure.rar

สำนักงานราชบัณฑิตยสภา. (2561). *พจนานุกรมศัพท์สถิติศาสตร์ ฉบับราชบัณฑิตยสภา* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำนักงานสถิติแห่งชาติ. *สาระน่ารู้ทางสถิติ.* สืบค้นเมื่อ 28 เมษายน 2562, จาก http://service.nso.go.th/nso/nsopublish/know/estat1_12.html

สำนักงานสถิติแห่งชาติ. *สำมะโน.* สืบค้นเมื่อ 5 มิถุนายน 2562, จาก http://service.nso.go.th/nso/nsopublish/census/census.html

สำนักทันตสาธารณสุข กรมอนามัย. *การเฝ้าระวังฟลูออไรด์ในน้ำบริโภค.* สืบค้นเมื่อ 1 สิงหาคม 2562, จาก http://dental2.anamai.moph.go.th/main.php?filename=fluoride_lab

Huff, D. (1993). *How to lie with statistics.* New York, NY: W. W. Norton & Company.

International Olympic Committee. *Winter Games.* Retrieved August 1, 2019, from https://www.olympic.org/winter-games

International Mathematical Olympiad. *55th IMO 2014: Country Results.* Retrieved March 20, 2019, from https://www.imo-official.org/year_country_r.aspx?year=2014

National Sleep Foundation. (2015, February 2). *National Sleep Foundation Recommends New Sleep Times.* Retrieved March 20, 2019, from https://www.sleepfoundation.org/press-release/national-sleep-foundation-recommends-new-sleep-times

Weiss, N. A. (2017). *Introductory Statistics* (10th ed). Essex, England: Pearson Education Limited.

ที่มาของข้อมูล

หน้า 14	(ตัวอย่างที่ 3)	http://www.nso.go.th/sites/2014/DocLib13/ด้านสังคม/สาขารายได้/เศรษฐกิจสังคมครัวเรือน/60/SocioSum60_First-6-Month.pdf สืบค้นเมื่อ 23 สิงหาคม 2560
หน้า 15	(แบบฝึกหัด 1.2 ข้อ 2)	http://suandusitpoll.dusit.ac.th/UPLOAD_FILES/POLL/2562/PS-2562-1549078433.pdf สืบค้นเมื่อ 20 มีนาคม 2562
หน้า 21	(แบบฝึกหัด 1.3 ข้อ 2)	http://www.nso.go.th/sites/2014/DocLib13/ด้านสังคม/สาขาสุขภาพ/พฤติกรรมการบริโภคอาหารของประชากร/การสำรวจพฤติกรรมการบริโภคอาหารของประชากร_2560/สรุปผลสำคัญ.pdf สืบค้นเมื่อ 24 พฤษภาคม 2561
หน้า 24	(ตัวอย่างของการใช้สถิติศาสตร์เชิงอนุมาน)	http://www.nso.go.th/sites/2014en/Survey/economics/Tourism%20and%20Sports/2016/Travel_behavior_2016_Full%20%20Report.pdf สืบค้นเมื่อ 26 พฤษภาคม 2560
หน้า 25	(แบบฝึกหัด 1.4 ข้อ 2)	http://www.nso.go.th/sites/2014/Documents/pop/2553/Results_report2553.pdf สืบค้นเมื่อ 30 เมษายน 2557

หน้า 26	(แบบฝึกหัด 1.4 ข้อ 4)	http://www.serviceapp.niets.or.th/onetmap สืบค้นเมื่อ 16 กรกฎาคม 2560
หน้า 27	(แบบฝึกหัด 1.4 ข้อ 6)	http://rh.anamai.moph.go.th/download/all_file/index/RH@2559.pdf สืบค้นเมื่อ 24 สิงหาคม 2560
หน้า 28	(แบบฝึกหัด 1.4 ข้อ 7)	http://www.nso.go.th/sites/2014/DocLib13/ด้านสังคม/สาขารายได้/การออมภาคครัวเรือน/ไตรมาส_3_2561.pdf สืบค้นเมื่อ 18 ธันวาคม 2561
หน้า 31	(แบบฝึกหัดท้ายบท ข้อ 6)	http://www.asianvolleyball.org/upfile/bulletin/files/201410/876894764_1412266044.pdf สืบค้นเมื่อ 18 ธันวาคม 2557
หน้า 32	(แบบฝึกหัดท้ายบท ข้อ 7)	https://www.th.undp.org/content/dam/thailand/docs/publications/nhdr14_en.pdf สืบค้นเมื่อ 3 กรกฎาคม 2558
หน้า 33	(แบบฝึกหัดท้ายบท ข้อ 8 2))	http://164.115.22.198/nso/phpview.php?folder=uploads/mnre/File/&partfile=Data_21030100_0102.xlsx สืบค้นเมื่อ 24 มกราคม 2562
หน้า 34	(แบบฝึกหัดท้ายบท ข้อ 8 4))	http://www.thairsc.com สืบค้นเมื่อ 3 มิถุนายน 2561
หน้า 49	(ตัวอย่างที่ 5)	https://www.doe.go.th/prd/assets/upload/files/overseas_th/07249dbf6f2e08c641d61ddc7fa40115.pdf สืบค้นเมื่อ 21 มิถุนายน 2557

หน้า 60	(แบบฝึกหัด 2.2 ข้อ 1)	http://www.nso.go.th/sites/2014/DocLib13/ด้านสังคม/สาขาแรงงาน/ภาวะการทำงานของประชากร/2562/Report_05-62.pdf สืบค้นเมื่อ 1 สิงหาคม 2562
หน้า 62	(แบบฝึกหัด 2.2 ข้อ 3)	https://www.etda.or.th/publishing-detail/thailand-internet-user-profile-2017.html สืบค้นเมื่อ 7 กรกฎาคม 2561
หน้า 68	(แบบฝึกหัดท้ายบท ข้อ 5)	https://ipricethailand.com/insights/mapofecommerce สืบค้นเมื่อ 5 มิถุนายน 2562
หน้า 97	(แผนภาพจุด)	https://www.bbc.com/sport/winter-olympics/medals/countries สืบค้นเมื่อ 4 สิงหาคม 2561
หน้า 113	(แบบฝึกหัด 3.2 ข้อ 3)	http://203.155.220.118/green-parks-admin/reports/report19.php?park_year=2562&rang1=2556&rang2=2562 สืบค้นเมื่อ 22 กันยายน 2562
หน้า 139	(ตัวอย่างที่ 18)	https://apps.fas.usda.gov/psdonline/circulars/Sugar.pdf สืบค้นเมื่อ 19 สิงหาคม 2562
หน้า 145	(ตัวอย่างที่ 21)	Wiesenfeld, L. P. (2001). The World Almanac and Book of Facts. Mahway, NJ: World Almanac Books.
หน้า 152	(แบบฝึกหัด 3.3.2 ข้อ 4)	http://www.thairsc.com/TH/stataccidentonfestival สืบค้นเมื่อ 26 กรกฎาคม 2559
หน้า 152	(แบบฝึกหัด 3.3.2 ข้อ 6)	http://service.nso.go.th/nso/web/statseries/tables/44000_Khon_Kaen/temp-46-58.xls สืบค้นเมื่อ 10 สิงหาคม 2559

หน้า 154 (ตัวอย่างที่ 23)	http://www.thairiceexporters.or.th/Press%20release/2018/TREA%20Press%20Release%20Thai%20Rice%20Situation%20&%20Trend%202018-31012018.pdf สืบค้นเมื่อ 28 สิงหาคม 2561
หน้า 159 (แบบฝึกหัด 3.3.3 ข้อ 1)	http://www.egat.co.th/index.php?option=com_content&view=article&id=452&Itemid=116 สืบค้นเมื่อ 28 เมษายน 2562
หน้า 161 (แบบฝึกหัด 3.3.3 ข้อ 3)	Wiesenfeld, L. P. (2001). The World Almanac and Book of Facts. Mahway, NJ: World Almanac Books.
หน้า 162 (แบบฝึกหัด 3.3.3 ข้อ 5)	http://service.nso.go.th/nso/nsopublish/themes/files/sumPay56.pdf สืบค้นเมื่อ 18 ธันวาคม 2557
หน้า 168 (แบบฝึกหัดท้ายบท ข้อ 3)	http://www.dnp.go.th/statistics/2556/ตาราง 31 น้ำฝนรายเดือน ปี 2554-2556.xls สืบค้นเมื่อ 30 เมษายน 2557
หน้า 184 (แบบฝึกหัดท้ายบท ข้อ 31)	http://www.dnp.go.th/statistics/2556/ตาราง 31 น้ำฝนรายเดือน ปี 2554-2556.xls สืบค้นเมื่อ 30 เมษายน 2557
หน้า 187 (แบบฝึกหัดท้ายบท ข้อ 35)	https://s3.amazonaws.com/happiness-report/2017/HR17.pdf สืบค้นเมื่อ 28 สิงหาคม 2560
หน้า 248 (แบบฝึกหัดท้ายบท ข้อ 4)	http://www.glo.or.th/main.php?filename=glo_lotto สืบค้นเมื่อ 17 มิถุนายน 2562

ที่มาของภาพ

หน้าสารบัญ, 1, 11	ดัดแปลงจาก © Dan Kernler/Wikimedia Commons/CC BY-SA 4.0
หน้า 24	Freepik.com
หน้า 76	Agaligo Coffee
หน้า 168	© user:Matthew Ferguson 57/Wikimedia Commons/CC BY-SA 4.0
หน้า 191	Crazy About Cafe

กราฟทั้งหมดในหนังสือเรียนเล่มนี้สร้างด้วยโปรแกรม GeoGebra

ภาคผนวก

ดัชนี

บทที่ 2

		หน้า
ความถี่	frequency	37
ฐานนิยม	mode	37
การแจกแจงความถี่	frequency distribution	40
ตารางความถี่	frequency table	40
ตารางความถี่จำแนกทางเดียว	one-way frequency table	40
ตารางความถี่จำแนกสองทาง	two-way frequency table	40
ความถี่สัมพัทธ์	relative frequency	41
ความถี่ร่วม	joint frequency	45
แผนภูมิรูปภาพ	pictogram	48
แผนภูมิรูปวงกลม	pie chart, circular chart	50
แผนภูมิแท่ง	bar chart	53
แผนภูมิแท่งเชิงเดี่ยว	simple bar chart	53
แผนภูมิแท่งพหุคูณ	multiple bar chart	55
แผนภูมิแท่งส่วนประกอบ	component bar chart	57

บทที่ 3

		หน้า
อันตรภาคชั้น	class interval	78
อันตรภาคชั้นเปิด	open-ended class interval	82
ฮิสโทแกรม	histogram	90
ขอบล่างของชั้น	lower class boundary	93
ขอบบนของชั้น	upper class boundary	93
แผนภาพจุด	dot plot	97
แผนภาพลำต้นและใบ	stem and leaf plot	99
แผนภาพกล่อง	box plot	102
ควอร์ไทล์	quartile	102, 153

การแจกแจงทวินาม	binomial distribution	213
การลองแบร์นูลลี	Bernoulli trial	214
เส้นโค้งความหนาแน่น	density curve	223
ฟังก์ชันความหนาแน่นความน่าจะเป็น	probability density function	223
การแจกแจงปกติ	normal distribution	225
เส้นโค้งปกติ	normal curve	226
การแจกแจงปกติมาตรฐาน	standard normal distribution	228
เส้นโค้งปกติมาตรฐาน	standard normal curve	228
ตัวแปรสุ่มปกติมาตรฐาน	standard normal random variable	228

บัญชีสัญลักษณ์

บทที่ 3

Q_i	ควอร์ไทล์ที่ i
μ	ค่าเฉลี่ยเลขคณิตของประชากร
$\overline{x}$	ค่าเฉลี่ยเลขคณิตของตัวอย่าง
$\sum_{i=1}^{n} x_i$	$x_1 + x_2 + x_3 + \cdots + x_n$
IQR	พิสัยระหว่างควอร์ไทล์
σ	ส่วนเบี่ยงเบนมาตรฐานของประชากร
s	ส่วนเบี่ยงเบนมาตรฐานของตัวอย่าง
σ^2	ความแปรปรวนของประชากร
s^2	ความแปรปรวนของตัวอย่าง
P_i	เปอร์เซ็นไทล์ที่ i

บทที่ 4

μ_X	ค่าคาดหมายของตัวแปรสุ่มไม่ต่อเนื่อง X
σ_X	ส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่มไม่ต่อเนื่อง X
σ_X^2	ความแปรปรวนของตัวแปรสุ่มไม่ต่อเนื่อง X
$X \sim B(n, p)$	การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงทวินามที่มี n และ p เป็นพารามิเตอร์
$X \sim N(\mu, \sigma^2)$	การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม X เป็นการแจกแจงปกติที่มี μ และ σ^2 เป็นพารามิเตอร์

คณะผู้จัดทำ

ที่ปรึกษา

ศ. ดร.ชูกิจ ลิมปิจำนงค์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

คณะผู้จัดทำหนังสือเรียน

นายประสาท สอ้านวงศ์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
รศ. ดร.สมพร สูตินันท์โอภาส	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
รศ. ดร.สิริพร ทิพย์คง	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวจินตนา อารยะรังสฤษฏ์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นายสุเทพ กิตติพิทักษ์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวจำเริญ เจียวหวาน	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
ดร.อลงกรณ์ ตั้งสงวนธรรม	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวปฐมาภรณ์ อวชัย	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวอัมริสา จันทนะศิริ	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นายพัฒนชัย รวิวรรณ	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวภิญญดา กลับแก้ว	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
ดร.ศศิวรรณ เมลืองนนท์	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
ดร.สุธารส นิลรอด	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
ดร.จิณณวัตร เจตน์จรุงกิจ	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
รศ. ดร.กมล บุษบา	มหาวิทยาลัยธรรมศาสตร์
ผศ. ดร.จุฑาภรณ์ สินสมบูรณ์ทอง	มหาวิทยาลัยเกษตรศาสตร์
รศ. ดร.ณัฐกาญจน์ ใจดี	จุฬาลงกรณ์มหาวิทยาลัย
ผศ. ดร.อนุภาพ สมบูรณ์สวัสดี	จุฬาลงกรณ์มหาวิทยาลัย
ผศ. ดร.วิฐรา พึ่งพาพงศ์	จุฬาลงกรณ์มหาวิทยาลัย

ดร.อัครินทร์ ไพบูลย์พานิช	จุฬาลงกรณ์มหาวิทยาลัย
ดร.นัท กุลวานิช	จุฬาลงกรณ์มหาวิทยาลัย
ผศ. ดร.จิราพรรณ สุนทรโชติ	จุฬาลงกรณ์มหาวิทยาลัย
ดร.เรวัต ถนัดกิจหิรัญ	จุฬาลงกรณ์มหาวิทยาลัย
ผศ. ดร.จุฑารัตน์ วรประทีป	มหาวิทยาลัยรามคำแหง
ดร.รชยา ศรีสุริฉัน	นักวิชาการอิสระ
ดร.พรฑิตา ทิวทัศน์	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

บรรณาธิการ

รศ. ดร.สุพล ดุรงค์วัฒนา	จุฬาลงกรณ์มหาวิทยาลัย

ฝ่ายสนับสนุนวิชาการ

รศ. ดร.วิชาญ ลิ่วกีรติยุตกุล	จุฬาลงกรณ์มหาวิทยาลัย
รศ. ดร.สัญญา มิตรเอม	มหาวิทยาลัยธรรมศาสตร์
ผศ. ดร.มุกดา แม้นมินทร์	ข้าราชการบำนาญ
ผศ. ดร.ปราณี นิลกรณ์	ข้าราชการบำนาญ
ดร.บำรุงศักดิ์ เผื่อนอารีย์	มหาวิทยาลัยบูรพา
นางนงนุช ผลทวี	โรงเรียนทับปุดวิทยา จังหวัดพังงา
นายชัยรัตน์ สุนทรประพี	นักวิชาการอิสระ
สาขาวิทยาศาสตร์มัธยมศึกษาตอนปลาย	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
ฝ่ายนวัตกรรมเพื่อการเรียนรู้	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
นางสาวปิยาภรณ์ ทองมาก	สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี

ออกแบบปก

บริษัท พิงค์ บลู แบล็ค แอนด์ ออเร้งจ์ จำกัด

ออกแบบรูปเล่ม

บริษัท เพอลังอิ พับลิชชิ่ง (ประเทศไทย) จำกัด

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี
กระทรวงศึกษาธิการ